

# LinearRegresssionGenderScopus2minimalRobustLoop.R

MikeT

Wed Jun 12 08:17:28 2019

```
library(car) #vif
library(psych) # Geometric.mean

##
## Attaching package: 'psych'

## The following object is masked from 'package:car':
##
##      logit

library(robustbase) #lmrob for linear regression in the presence of
heteroscedasticity https://stackoverflow.com/questions/23893103/increase-maximum-number-of-iterations-in-lmrob

##
## Attaching package: 'robustbase'

## The following object is masked from 'package:psych':
##
##      cushny

#library(plyr) #for count
library(expss) #for countif

## Loading required package: magrittr

## Loading required package: htmlTable

##
## Attaching package: 'expss'

## The following object is masked from 'package:car':
##
##      recode

outlier_threshold = 2.5
options(digits=4)
#FirstCode <- 1000
#LastCode <- 3616 #3616
#row.names <- c(FirstCode:LastCode)
#FieldCount <- LastCode - FirstCode + 1

FieldCodeSet <-
c(1000,1100,1101,1102,1103,1104,1105,1106,1107,1108,1109,1110,1111,1200,1201,
```

```

1202,1203,1204,1205,1206,1207,1208,1209,1210,1211,1212,1213,1300,1301,1302,13
03,1304,1305,1306,1307,1308,1309,1310,1311,1312,1313,1314,1315,1400,1401,1402
,1403,1404,1405,1406,1407,1408,1409,1410,1500,1501,1502,1503,1504,1505,1506,1
507,1508,1600,1601,1602,1603,1604,1605,1606,1607,1700,1701,1702,1703,1704,170
5,1706,1707,1708,1709,1710,1711,1712,1800,1801,1802,1803,1804,1900,1901,1902,
1903,1904,1905,1906,1907,1908,1909,1910,1911,1912,1913,2000,2001,2002,2003,21
00,2101,2102,2103,2104,2105,2200,2201,2202,2203,2204,2205,2206,2207,2208,2209
,2210,2211,2212,2213,2214,2215,2216,2300,2301,2302,2303,2304,2305,2306,2307,2
308,2309,2310,2311,2312,2400,2401,2402,2403,2404,2405,2406,2500,2501,2502,250
3,2504,2505,2506,2507,2508,2600,2601,2602,2603,2604,2605,2606,2607,2608,2609,
2610,2611,2612,2613,2614,2700,2701,2702,2703,2704,2705,2706,2707,2708,2709,27
10,2711,2712,2713,2714,2715,2716,2717,2718,2719,2720,2721,2722,2723,2724,2725
,2726,2727,2728,2729,2730,2731,2732,2733,2734,2735,2736,2737,2738,2739,2740,2
741,2742,2743,2744,2745,2746,2747,2748,2800,2801,2802,2803,2804,2805,2806,280
7,2808,2809,2900,2901,2902,2903,2904,2905,2906,2907,2908,2909,2910,2911,2912,
2913,2914,2915,2916,2917,2918,2919,2920,2921,2922,2923,3000,3001,3002,3003,30
04,3005,3100,3101,3102,3103,3104,3105,3106,3107,3108,3109,3110,3200,3201,3202
,3203,3204,3205,3206,3207,3300,3301,3302,3303,3304,3305,3306,3307,3308,3309,3
310,3311,3312,3313,3314,3315,3316,3317,3318,3319,3320,3321,3322,3400,3401,340
2,3403,3404,3500,3501,3502,3503,3504,3505,3506,3600,3601,3602,3603,3604,3605,
3606,3607,3608,3609,3610,3611,3612,3613,3614,3615,3616)
#FieldCodeSet <- 2700:2748
FieldNeedsExtraTime <- FieldCodeSet #Allow the extra time for all fields,
just in case
#FieldNeedsExtraTime <- c(1311, 3100, 2700, 2709, 2714, 2719,2722, 2734,
2744, 2747, 3104, 3107, 1701, 1706, 3302,3303,3309, 3320, 3322) #Fields known
to need extra time
row.names <- FieldCodeSet
FieldCount <- length(FieldCodeSet)
FirstYearForCitationAnalysis <- 1996
LastYearForCitationAnalysis <- 2012
MaxAuthorsToCountInRegression <- 5;
MaxCountriesAllowed <- 1; #set to 0 to skip
#One country only unless comment below.
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\All 1996-2018 332
fields was Jamaica\\cov3\\ScopusFind Jamaica"; Country <- "All"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\Australia 1996-
2018 331 fields\\cov3\\ScopusFind Australia"; Country <- "Aus"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\Canada 1996-2018
331 fields\\cov3\\ScopusFind Canada"; Country <- "Ca"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\China 1996-2018
327 fields\\cov3\\ScopusFind China"; Country <- "CN"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\Germany 1996-2018
330 fields\\cov3\\ScopusFind Germany"; Country <- "De"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\Ireland 1996-2018
329 fields\\cov3\\ScopusFind Ireland"; Country <- "Ie"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\Spain 1996-2018
329 fields\\cov3\\ScopusFind Spain"; Country <- "Es"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\NZ 1996-2018 327
fields\\cov3\\ScopusFind New Zealand"; Country <- "Nz"

```

```

RootFileName <- "E:\\data\\Scopus\\All fields regression\\UK gender 1996-2018
330 fields\\cov3\\ScopusFind United Kingdom"; Country <- "UK"
#RootFileName <- "E:\\data\\Scopus\\All fields regression\\USA 1996-2018 331
fields\\cov3\\United States "; Country <- "Usa"
column.names <- c("Articles",
"FirstF","FirstM","FirstP","LastF","LastM","LastP")
GenderTeamSize <- array(integer(7*FieldCount),dim = c(FieldCount, 7),
dimnames <- list(row.names, column.names))
column.names <- c("FFA1","FLA1","2","3","4","5+","FFA2",
"FLA2","FFA3","FLA4")
RegCoef <- array(integer(10*FieldCount),dim = c(FieldCount, 10), dimnames <-
list(row.names, column.names))
column.names <- c("FFA1p","FLA1p","2p","3p","4p","5+p","FFA2p",
"FLA2p","FFA3p","FLA4p")
RegP <- array(integer(10*FieldCount),dim = c(FieldCount, 10), dimnames <-
list(row.names, column.names))
RegStar <- array(character(10*FieldCount),dim = c(FieldCount, 10), dimnames
<- list(row.names, column.names))
column.names <- c("1Ctry","Gend 1st nth","Fem1 96","Fem1 18","Chg1","FemN
96","FemN 18", "ChgN")
BasicStats <- array(integer(8*FieldCount),dim = c(FieldCount, 8), dimnames <-
list(row.names, column.names))
column.names <- c("1CtyMFto14","Fem1All","Fem1 96","Fem1
14","Chg1","FemNAll","FemN96","FemN14", "ChgN", "1CtyMF96", "1CtyMF14")
BasicStats2 <- array(integer(11*FieldCount),dim = c(FieldCount, 11), dimnames
<- list(row.names, column.names))
column.names <- c("CitationSet","Female1st","FemaleLast")
GenderAnalysed <- array(integer(3*FieldCount),dim = c(FieldCount, 3),
dimnames <- list(row.names, column.names))

for (i in 1:FieldCount) {
  for (j in 1:7) {
    GenderTeamSize[i,j] <- NA; BasicStats[i,j]<-NA; BasicStats2[i,j]<-NA;
RegCoef[i,j]<-NA ; RegP[i,j]<-NA
  }
  BasicStats[i,8]<-NA; BasicStats2[i,8]<-NA; RegCoef[i,8]<-NA ; RegP[i,8]<-NA
  BasicStats2[i,9]<-NA; RegCoef[i,9]<-NA ; RegP[i,9]<-NA
  BasicStats2[i,10]<-NA;RegCoef[i,10]<-NA ; RegP[i,10]<-NA
  BasicStats2[i,11]<-NA
}

FieldCount = 0
#iCode <- 3302 #test
for (iCode in FieldCodeSet) {
  FieldCount <- FieldCount + 1 #
  SampleFile <- paste(RootFileName, iCode, "_cov.txt", sep="");
  BasicStats[FieldCount,1] <- 0; BasicStats2[FieldCount,1] <- 0;
  if (file.exists(SampleFile)) {
    tryCatch({
      print(""); print("")
    })
  }
}

```

```

print("#####")
print(paste("Analysis of AJSC",iCode))
print("#####")
AllScopusData <- read.table(file=SampleFile, head=TRUE, sep = "\t")
names(AllScopusData)[3] <- "NLCS"
names(AllScopusData)[4] <- "Year"
names(AllScopusData)[5] <- "OneField"
names(AllScopusData)[6] <- "Fields"
#AllScopusData <- AllScopusData[AllScopusData$Year!=2004,] #Uncomment
this to remove year with high outliers if code below reveals a problem
if (MaxCountriesAllowed>0) AllScopusData <-
AllScopusData[AllScopusData$UniqueCountries <= MaxCountriesAllowed,] #Comment
out to include non-us authors after 1st
AllScopusDataOlder <- AllScopusData[AllScopusData$Year <=
LastYearForCitationAnalysis,] #2015 Ensure citation window of at least 3
years 2013 for 5 years
AllScopusDataOlder$Year <- factor(AllScopusDataOlder$Year) #Treat each
year separately
AllScopusDataOlderFirstGendered <-
AllScopusDataOlder[AllScopusDataOlder$FirstAuthorFemale > -1,] #Female 1,
Male 0, Unknown -1
AllScopusDataOlderFirstGendered$FirstAuthorFemale <-
factor(AllScopusDataOlderFirstGendered$FirstAuthorFemale)
AllScopusDataOlderFirstLastGendered <-
AllScopusDataOlderFirstGendered[AllScopusDataOlderFirstGendered$LastAuthorFemale > -1,] #Female 1, Male 0, Unknown -1
AllScopusDataOlderFirstLastGendered$LastAuthorFemale <-
factor(AllScopusDataOlderFirstLastGendered$LastAuthorFemale) #Female 1, Male
0, Unknown -1
AllYearsBothGendered <- AllScopusData[AllScopusData$FirstAuthorFemale >
-1,]
AllYearsBothGendered <-
AllYearsBothGendered[AllYearsBothGendered$LastAuthorFemale > -1,]
AllYearsBothGendered1996 <-
AllYearsBothGendered[AllYearsBothGendered$Year == 1996,]
AllYearsBothGenderedLast <-
AllYearsBothGendered[AllYearsBothGendered$Year ==
LastYearForCitationAnalysis,]
AllYearsBothGendered2018 <-
AllYearsBothGendered[AllYearsBothGendered$Year == 2018,]
BasicStats[FieldCount,1] <- nrow(AllScopusData)
BasicStats[FieldCount,2] <- nrow(AllYearsBothGendered) /
nrow(AllScopusData) * 100 #Percentage
BasicStats[FieldCount,3] <- count_if(1,
AllYearsBothGendered1996$FirstAuthorFemale)/ nrow(AllYearsBothGendered1996) *
100
BasicStats[FieldCount,4] <- count_if(1,
AllYearsBothGendered2018$FirstAuthorFemale)/ nrow(AllYearsBothGendered2018) *
100
BasicStats[FieldCount,5] <- BasicStats[FieldCount,4] -

```



```

BasicStats[FieldCount,3]
  BasicStats[FieldCount,6] <- count_if(1,
AllYearsBothGendered1996$LastAuthorFemale)/ nrow(AllYearsBothGendered1996) *
100
  BasicStats[FieldCount,7] <- count_if(1,
AllYearsBothGendered2018$LastAuthorFemale)/ nrow(AllYearsBothGendered2018) *
100
  BasicStats[FieldCount,8] <- BasicStats[FieldCount,7] -
BasicStats[FieldCount,6]
  ##c("US gendered96to14","Fem1 All","Fem1 96","Fem1 14","Chg","FemN
All","FemN 96","FemN 14", "Chg")
  BasicStats2[FieldCount,1] <- nrow(AllScopusDataOlderFirstLastGendered)
  BasicStats2[FieldCount,2] <- count_if(1,
AllScopusDataOlderFirstLastGendered$FirstAuthorFemale) /
nrow(AllScopusDataOlderFirstLastGendered) * 100 #Percentage
  BasicStats2[FieldCount,3] <- count_if(1,
AllYearsBothGendered1996$FirstAuthorFemale)/ nrow(AllYearsBothGendered1996) *
100
  BasicStats2[FieldCount,4] <- count_if(1,
AllYearsBothGenderedLast$FirstAuthorFemale)/ nrow(AllYearsBothGenderedLast) *
100
  BasicStats2[FieldCount,5] <- BasicStats2[FieldCount,4] -
BasicStats2[FieldCount,3]
  BasicStats2[FieldCount,6] <- count_if(1,
AllScopusDataOlderFirstLastGendered$LastAuthorFemale) /
nrow(AllScopusDataOlderFirstLastGendered) * 100
  BasicStats2[FieldCount,7] <- count_if(1,
AllYearsBothGendered1996$LastAuthorFemale)/ nrow(AllYearsBothGendered1996) *
100
  BasicStats2[FieldCount,8] <- count_if(1,
AllYearsBothGenderedLast$LastAuthorFemale)/ nrow(AllYearsBothGenderedLast) *
100
  BasicStats2[FieldCount,9] <- BasicStats2[FieldCount,8] -
BasicStats2[FieldCount,7]
  BasicStats2[FieldCount,10] <- nrow(AllYearsBothGendered1996)
  BasicStats2[FieldCount,11] <- nrow(AllYearsBothGenderedLast)
  #print("MNLCS for all years [All, first gendered, first & last
gendered], just to check nothing is odd")
  #print(tapply(AllScopusDataOlder$NLCS, AllScopusDataOlder$Year, mean))
  #print(tapply(AllScopusDataOlderFirstGendered$NLCS,
AllScopusDataOlderFirstGendered$Year, mean))
  #print(tapply(AllScopusDataOlderFirstLastGendered$NLCS,
AllScopusDataOlderFirstLastGendered$Year, mean))

  print("Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]")
  print(table(AllScopusDataOlder$Year))
  print(table(AllScopusDataOlderFirstGendered$Year))
  print(table(AllScopusDataOlderFirstLastGendered$Year))
}, error = function(e) return("failed narrow field data entry and basic

```

```

processing"))
  tryCatch({
    #####
    print("Heteroscedasticity checks, confirming that there are problems
with these")
    #####
    #Check for outliers caused by field normalisation - should be evident
in heteroskedacity caused by individual years
    print(bartlett.test(NLCS~Year, data=AllScopusDataOlderFirstGendered))
#Homogeneity of Variances test. Big fail is OK because older years have
greater variability
    YearLm <- lm(NLCS~Year, data=AllScopusDataOlderFirstGendered)
    plot(YearLm, which = 1) #Residuals vs. fitted. Check that width of
residual bar is not huge for a few years
    AllScopusDataOlderFirstGendered$YMresiduals <- resid(YearLm)
    YearGenderLm <- lm(YMresiduals~FirstAuthorFemale,
data=AllScopusDataOlderFirstGendered)
    print(bartlett.test(YMresiduals~FirstAuthorFemale,
data=AllScopusDataOlderFirstGendered)) #Homogeneity of Variances test - this
is the key test - should not fail by much, but might becuae older years have
fewer females and higher variance
    plot(YearGenderLm, which = 1) #Residuals vs. fitted. Check that width
of residual bar does not vary too much for homoscedasticity; look out for
cone shape
    #If problems here, check the outliers in regression 1
  }, error = function(e) return("failed Heteroscedasticity checks"))
  tryCatch({
    #####
    # Test for different team sizes for male and female authors - first and
last gendered
    #####
    #First author gender 2018
    GenderAuthors2018 <-
split(AllYearsBothGendered2018$UniqueAuthors,AllYearsBothGendered2018$FirstAu
thorFemale)
    FemaleTeamSizes2018 <- as.numeric(GenderAuthors2018$'1')
    MaleTeamSizes2018 <- as.numeric(GenderAuthors2018$'0')
    GenderAnalysed[FieldCount,1] <- nrow(AllYearsBothGendered2018)
    GenderAnalysed[FieldCount,2] <- length(FemaleTeamSizes2018)
    print(paste("Female first author team size 2018 geometric mean:",
geometric.mean(FemaleTeamSizes2018)))
    print(paste("Male first author team size 2018 geometric mean:",
geometric.mean(MaleTeamSizes2018)))
    wilc<- wilcox.test(FemaleTeamSizes2018, MaleTeamSizes2018, alternative
= "two.sided")
    print(wilc)
    GenderTeamSize[FieldCount,1] <- nrow(AllYearsBothGendered2018)
    GenderTeamSize[FieldCount,2] <- geometric.mean(FemaleTeamSizes2018)
    GenderTeamSize[FieldCount,3] <- geometric.mean(MaleTeamSizes2018)
    GenderTeamSize[FieldCount,4] <- wilc$p.value
  })
}

```

```

#Last author gender 2018
GenderAuthors2018 <-
split(AllYearsBothGendered2018$UniqueAuthors,AllYearsBothGendered2018$LastAuthorFemale)
  FemaleTeamSizes2018 <- as.numeric(GenderAuthors2018$'1')
  MaleTeamSizes2018 <- as.numeric(GenderAuthors2018$'0')
  GenderAnalysed[FieldCount,3] <- length(FemaleTeamSizes2018)
  print(paste("Female last author team size 2018 geometric mean:",
geometric.mean(FemaleTeamSizes2018)))
  print(paste("Male last author team size 2018 geometric mean:",
geometric.mean(MaleTeamSizes2018)))
  wilc<- wilcox.test(FemaleTeamSizes2018, MaleTeamSizes2018, alternative
= "two.sided")
  print(wilc)
  GenderTeamSize[FieldCount,5] <- geometric.mean(FemaleTeamSizes2018)
  GenderTeamSize[FieldCount,6] <- geometric.mean(MaleTeamSizes2018)
  GenderTeamSize[FieldCount,7] <- wilc$p.value
}, error = function(e) return("Failed team size tests"))
tryCatch({
  #Use 5 for 5+ authors
  for (i in 1:nrow(AllScopusDataOlderFirstLastGendered)) {
    AllScopusDataOlderFirstLastGendered$UniqueAuthors[i] <-
min(MaxAuthorsToCountInRegression,AllScopusDataOlderFirstLastGendered$UniqueAuthors[i])
  }

AllScopusDataOlderFirstLastGendered$UniqueAuthors=factor(AllScopusDataOlderFirstLastGendered$UniqueAuthors) #Don't assume any team size influence formula
#####
  print("Regression 1: First author gender, last author gender, team
size, Year as factors")
#####
  if (iCode %in% FieldNeedsExtraTime) {
    FirstLastAuthorTeamLmrob <-
lmrob(NLCS~FirstAuthorFemale+LastAuthorFemale+UniqueAuthors+Year,
data=AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf, k.max = 1000))
  } else {
    FirstLastAuthorTeamLmrob <-
lmrob(NLCS~FirstAuthorFemale+LastAuthorFemale+UniqueAuthors+Year,
data=AllScopusDataOlderFirstLastGendered)
  }
  print(vif(FirstLastAuthorTeamLmrob)) # check under 5
  hist(resid(FirstLastAuthorTeamLmrob),main='Residuals from first and
last author and team size',xlab='Standardised Residuals',ylab='Frequency')
#Normality not needed but useful - check for outliers - if too many outside
+/- 1.96 then possible field normalisation problem
  AllScopusDataOlderFirstLastGendered$residuals <-
resid(FirstLastAuthorTeamLmrob)
  Outliers <- AllScopusDataOlderFirstLastGendered[

```

```

abs(AllScopusDataOlderFirstLastGendered$residuals) > outlier_threshold, ]
  Outliers <- Outliers[c(1,3,4,5,6,16)]
  Outliers$ScopusId <- gsub('scopus_id:', '', Outliers$ScopusId)
  print(paste("List of ", nrow(Outliers), "outliers with residuals above
", outlier_threshold))
  print(Outliers)
  sum <- summary(FirstLastAuthorTeamLmrob)
  print(sum)
  RegCoef[FieldCount,1] <- sum$coefficients[2]; RegP[FieldCount,1] <-
sum$coefficients[2,4];
  RegCoef[FieldCount,2] <- sum$coefficients[3]; RegP[FieldCount,2] <-
sum$coefficients[3,4];
  RegCoef[FieldCount,3] <- sum$coefficients[4]; RegP[FieldCount,3] <-
sum$coefficients[4,4];
  RegCoef[FieldCount,4] <- sum$coefficients[5]; RegP[FieldCount,4] <-
sum$coefficients[5,4];
  RegCoef[FieldCount,5] <- sum$coefficients[6]; RegP[FieldCount,5] <-
sum$coefficients[6,4];
  RegCoef[FieldCount,6] <- sum$coefficients[7]; RegP[FieldCount,6] <-
sum$coefficients[7,4];
}, error = function(e) return("Failed regression 1"))
tryCatch({
  #####
  print("Regression 2: First author gender, Last author gender, Year as
factors")
  #####
  #First author, last author regression
  if (iCode %in% FieldNeedsExtraTime) {
    FirstLastAuthorLmrob <-
lmrob(NLCS~FirstAuthorFemale+LastAuthorFemale+Year,
data=AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf, k.max = 1000))
  } else {
    FirstLastAuthorLmrob <-
lmrob(NLCS~FirstAuthorFemale+LastAuthorFemale+Year,
data=AllScopusDataOlderFirstLastGendered)
  }
  print(vif(FirstLastAuthorLmrob)) # check under 5
  hist(resid(FirstLastAuthorLmrob),main='Residuals from first and last
author',xlab='Standardised Residuals',ylab='Frequency') #Normality not needed
but useful - check for outliers - if too many outside +/- 1.96 then possible
field normalisation problem
  AllScopusDataOlderFirstLastGendered$residuals <-
resid(FirstLastAuthorLmrob)
  Outliers <- AllScopusDataOlderFirstLastGendered[
abs(AllScopusDataOlderFirstLastGendered$residuals) > outlier_threshold, ]
  Outliers <- Outliers[c(1,3,4,5,6,16)]
  Outliers$ScopusId <- gsub('scopus_id:', '', Outliers$ScopusId)
  print(paste("List of ", nrow(Outliers), "outliers with residuals above
", outlier_threshold))

```

```

    print(Outliers)
    sum <- summary(FirstLastAuthorLmrob)
    print(sum) #Robust to heteroscedacity
    RegCoef[FieldCount,7] <- sum$coefficients[2]; RegP[FieldCount,7] <-
sum$coefficients[2,4];
    RegCoef[FieldCount,8] <- sum$coefficients[3]; RegP[FieldCount,8] <-
sum$coefficients[3,4];
  }, error = function(e) return("Failed regression 4"))
  tryCatch({
    #####
    print("Regression 3: First author gender, Year as factors")
    #####
    #First author regression
    if (iCode %in% FieldNeedsExtraTime) {
      FirstAuthorLmrob <- lmrob(NLCS~FirstAuthorFemale+Year,
data=AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf, k.max = 1000))
    } else {
      FirstAuthorLmrob <- lmrob(NLCS~FirstAuthorFemale+Year,
data=AllScopusDataOlderFirstLastGendered)
    }
    print(vif(FirstAuthorLmrob)) # check under 5
    hist(resid(FirstAuthorLmrob),main='Residuals from first
author',xlab='Standardised Residuals',ylab='Frequency') #Normality not needed
but useful - check for outliers - if too many outside +/- 1.96 then possible
field normalisation problem
    AllScopusDataOlderFirstLastGendered$residuals <-
resid(FirstLastAuthorLmrob)
    Outliers <- AllScopusDataOlderFirstLastGendered[
abs(AllScopusDataOlderFirstLastGendered$residuals) > outlier_threshold, ]
    Outliers <- Outliers[c(1,3,4,5,6,16)]
    Outliers$ScopusId <- gsub('scopus_id:', '', Outliers$ScopusId)
    print(paste("List of ", nrow(Outliers), "outliers with residuals above
", outlier_threshold))
    print(Outliers)
    sum <- summary(FirstAuthorLmrob)
    print(sum) #Robust to heteroscedacity
    RegCoef[FieldCount,9] <- sum$coefficients[2]; RegP[FieldCount,9] <-
sum$coefficients[2,4];
  }, error = function(e) return("Failed regression 3"))
  tryCatch({
    #####
    print("Regression 4: Last author gender, Year as factors")
    #####
    if (iCode %in% FieldNeedsExtraTime) {
      LastAuthorLmrob <- lmrob(NLCS~LastAuthorFemale+Year,
data=AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf, k.max = 1000))
    } else {
      LastAuthorLmrob <- lmrob(NLCS~LastAuthorFemale+Year,

```

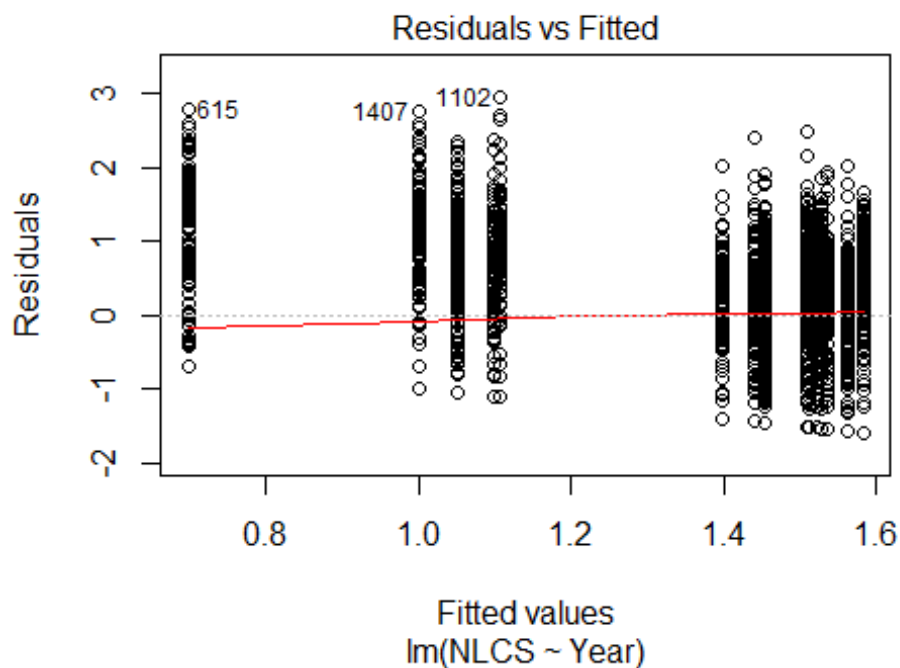
```

data=AllScopusDataOlderFirstLastGendered)
}
print(vif(LastAuthorLmrob)) # check under 5
hist(resid(FirstLastAuthorLmrob),main='Residuals from last
author',xlab='Standardised Residuals',ylab='Frequency') #Normality not needed
but useful - check for outliers - if too many outside +/- 1.96 then possible
field normalisation problem
AllScopusDataOlderFirstLastGendered$residuals <-
resid(FirstLastAuthorLmrob)
Outliers <- AllScopusDataOlderFirstLastGendered[
abs(AllScopusDataOlderFirstLastGendered$residuals) > outlier_threshold, ]
Outliers <- Outliers[c(1,3,4,5,6,16)]
Outliers$ScopusId <- gsub('scopus_id:', '', Outliers$ScopusId)
print(paste("List of ", nrow(Outliers), "outliers with residuals above
", outlier_threshold))
print(Outliers)
sum <- summary(LastAuthorLmrob) #Robust to heteroscedacity
print(sum) #Robust to heteroscedacity
#column.names <- c("FFA1", "FLA1", "2", "3", "4", "5+", "FFA2",
"FLA2", "FFA3", "FLA4")
#RegCoef <- array(integer(11*FieldCount),dim = c(FieldCount, 11),
dimnames <- list(row.names, column.names))
RegCoef[FieldCount,10] <- sum$coefficients[2]; RegP[FieldCount,10] <-
sum$coefficients[2,4];
}, error = function(e) return("Failed regression 4"))
tryCatch({
print(paste("Sample size for the above analysis: ",
nrow(AllScopusDataOlderFirstLastGendered)))
}, error = function(e) return("failed sample size"))
}
}

## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1000"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 301 394 322 328 320 316 292 197 230 245 220 260 240 228 266
## 2011 2012
## 281 313
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 183 274 206 227 195 159 204 138 159 183 161 199 175 187 220
## 2011 2012
## 237 244
##

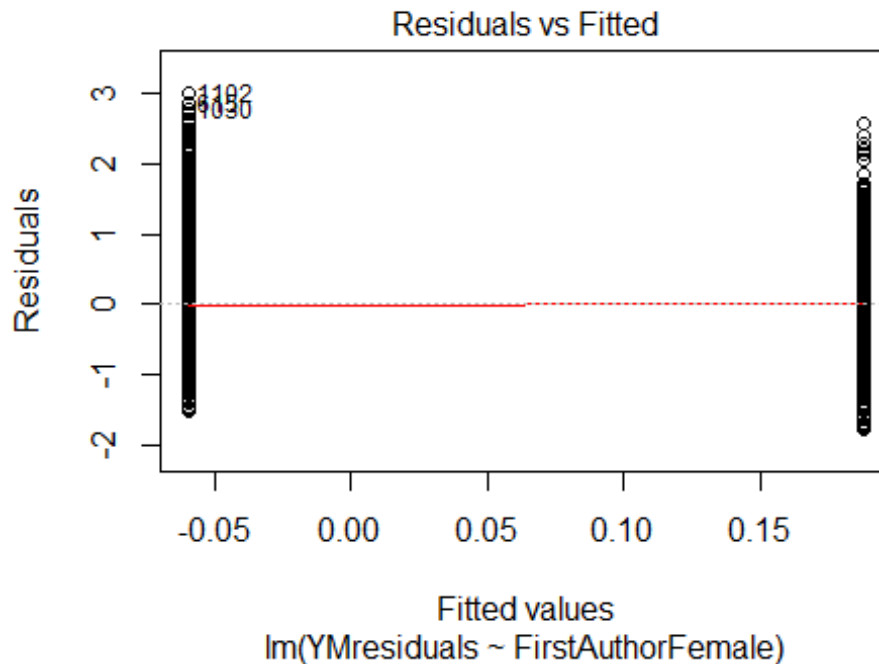
```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 175 265 197 213 182 148 180 126 141 159 143 180 158 168 199
## 2011 2012
## 215 225
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 190, df = 16, p-value <2e-16
```



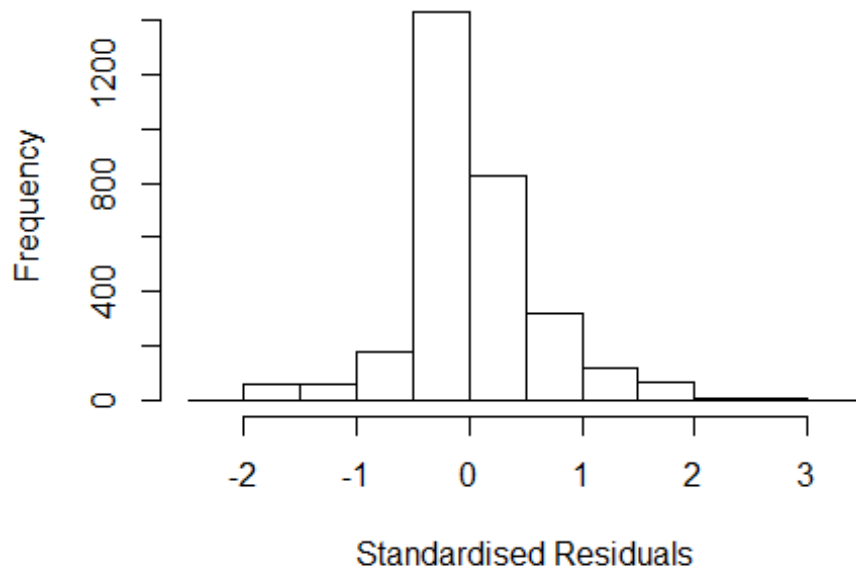
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 19, df = 1, p-value = 1e-05
```





```
## [1] "Female first author team size 2018 geometric mean: 4.18524399957648"
## [1] "Male first author team size 2018 geometric mean: 3.6111561090916"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 45000, p-value = 0.008
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.81273581811327"
## [1] "Male last author team size 2018 geometric mean: 3.81365504593466"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 31000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.413 1          1.189
## LastAuthorFemale  1.361 1          1.167
## UniqueAuthors     2.249 4          1.107
## Year              2.257 16         1.026
```

## Residuals from first and last author and team size



```
## [1] "List of 9 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 242   0029935458 2.742 1996    1000     2    2.610
## 305   0030071902 2.829 1996    1000     2    2.697
## 752   0030934532 3.116 1997    1000     1    3.015
## 753   0030939391 3.026 1997    1000     2    2.925
## 805   0031030138 2.727 1997    1000     2    2.626
## 1014  0032483336 2.797 1998    1000     2    2.608
## 1492  0033527009 2.908 1999    1000     1    2.716
## 2761  0037171855 2.579 2002    1000     2    2.514
## 4188  33745511017 2.915 2006    1000     1    2.817
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.052 -0.200 -0.059  0.312  3.015
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.13227    0.03101   4.27 2.1e-05 ***
## FirstAuthorFemale1 0.00505    0.02101   0.24  0.8102
```

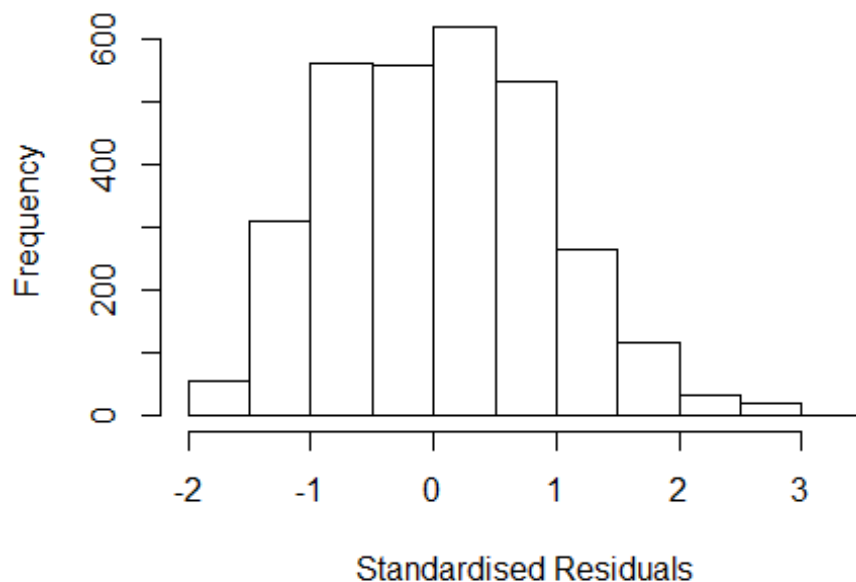
```

## LastAuthorFemale1 -0.04442 0.02481 -1.79 0.0735 .
## UniqueAuthors2 1.57801 0.04372 36.09 < 2e-16 ***
## UniqueAuthors3 1.64417 0.03297 49.86 < 2e-16 ***
## UniqueAuthors4 1.71116 0.03261 52.47 < 2e-16 ***
## UniqueAuthors5 1.78331 0.02508 71.09 < 2e-16 ***
## Year1997 -0.03140 0.03410 -0.92 0.3573
## Year1998 0.05701 0.03781 1.51 0.1317
## Year1999 0.05972 0.03983 1.50 0.1339
## Year2000 -0.07328 0.03730 -1.96 0.0495 *
## Year2001 0.05691 0.04168 1.37 0.1722
## Year2002 -0.06754 0.05167 -1.31 0.1912
## Year2003 -0.12527 0.05325 -2.35 0.0187 *
## Year2004 -0.03708 0.05672 -0.65 0.5133
## Year2005 -0.00239 0.04665 -0.05 0.9592
## Year2006 -0.03408 0.04656 -0.73 0.4642
## Year2007 0.00408 0.04731 0.09 0.9313
## Year2008 0.10609 0.04443 2.39 0.0170 *
## Year2009 0.06734 0.04745 1.42 0.1559
## Year2010 0.11930 0.04499 2.65 0.0081 **
## Year2011 0.27081 0.04866 5.56 2.9e-08 ***
## Year2012 0.23238 0.04963 4.68 3.0e-06 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared: 0.819, Adjusted R-squared: 0.817
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 88 observations
c(125,126,134,150,151,158,175,272,358,415,416,430,449,454,459,511,546,554,557
,564,575,577,584,620,629,642,653,695,699,704,717,726,754,779,786,804,807,828,
829,948,1010,1012,1028,1044,1066,1083,1174,1297,1360,1388,1399,1418,1435,1601
,1739,1780,1849,1997,2016,2108,2256,2257,2342,2391,2402,2462,2469,2503,2516,2
534,2581,2640,2648,2660,2661,2733,2765,2781,2782,2807,2836,2847,2854,2940,304
6,3047,3049,3053)
## are outliers with |weight| <= 2.5e-05 ( < 3.3e-05);
## 184 weights are ~= 1. The remaining 2802 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0001 0.8200 0.9560 0.8460 0.9870 0.9990
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 3.25e-05 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 1000 200
## trace.lev mts compute.rd

```

```
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.124 1          1.060
## LastAuthorFemale 1.066 1          1.032
## Year              1.112 16          1.003
```

### Residuals from first and last author



```
## [1] "List of 19 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 604 0030669546 3.039 1997 1000 2 2.684
## 664 0030795741 3.225 1997 1000 2 2.870
## 684 0030986236 2.880 1997 1000 1 2.525
## 691 0030878110 2.996 1997 1000 1 2.641
## 703 0030992349 3.279 1997 1000 2 2.924
## 705 0030989227 2.636 1997 1000 2 2.513
## 752 0030934532 3.116 1997 1000 1 2.761
## 753 0030939391 3.026 1997 1000 2 2.671
## 803 0031035644 2.963 1997 1000 2 2.608
## 1030 0032572722 3.822 1998 1000 2 2.949
## 1035 0032552114 3.760 1998 1000 2 2.887
## 1080 0032474853 3.416 1998 1000 2 2.543
```

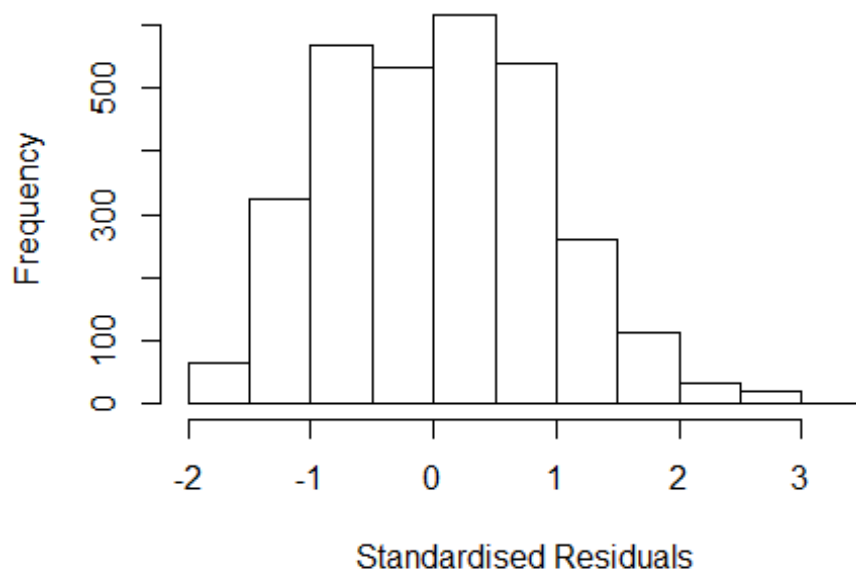
```

## 1102 0032511040 4.051 1998      1000      2      3.178
## 1407 0033607504 3.751 1999      1000      2      2.765
## 1434 0033615491 3.300 1999      1000      1      2.671
## 1473 0033539183 3.596 1999      1000      2      2.610
## 1534 0033542448 3.153 1999      1000      2      2.524
## 1556 0033587146 3.526 1999      1000      1      2.897
## 1800 0034676457 3.392 2000      1000      2      2.716
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8095 -0.6384  0.0446  0.6774  3.1779
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9794     0.1087   9.01 < 2e-16 ***
## FirstAuthorFemale1  0.3563     0.0382   9.34 < 2e-16 ***
## LastAuthorFemale1 -0.2319     0.0565  -4.10 4.2e-05 ***
## Year1997          -0.6245     0.1179  -5.30 1.3e-07 ***
## Year1998          -0.1062     0.1561  -0.68 0.49612
## Year1999          -0.3500     0.1478  -2.37 0.01793 *
## Year2000          -0.0717     0.1384  -0.52 0.60445
## Year2001          -0.0943     0.1570  -0.60 0.54813
## Year2002           0.4017     0.1179   3.41 0.00066 ***
## Year2003           0.3409     0.1309   2.60 0.00925 **
## Year2004           0.5128     0.1245   4.12 3.9e-05 ***
## Year2005           0.4121     0.1259   3.27 0.00107 **
## Year2006           0.4730     0.1245   3.80 0.00015 ***
## Year2007           0.5129     0.1278   4.01 6.2e-05 ***
## Year2008           0.5061     0.1322   3.83 0.00013 ***
## Year2009           0.5056     0.1314   3.85 0.00012 ***
## Year2010           0.4320     0.1367   3.16 0.00159 **
## Year2011           0.4738     0.1436   3.30 0.00098 ***
## Year2012           0.6130     0.1319   4.65 3.5e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.69
## Multiple R-squared:  0.189, Adjusted R-squared:  0.184
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 164 weights are ~= 1. The remaining 2910 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0011 0.8030 0.9050 0.8540 0.9760 0.9990
## Algorithmic parameters:

```

```
##          tuning.chi          bb          tuning.psi          refine.tol
##          1.55e+00          5.00e-01          4.69e+00          1.00e-07
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          3.25e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample          max.it          best.r.s          k.fast.s          k.max          maxit.scale
##          500          50          2          1          1000          200
## trace.lev          mts          compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.086 1          1.042
## Year          1.086 16          1.003
```

### Residuals from first author



```
## [1] "List of 19 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 604 0030669546 3.039 1997 1000 2 2.684
## 664 0030795741 3.225 1997 1000 2 2.870
## 684 0030986236 2.880 1997 1000 1 2.525
## 691 0030878110 2.996 1997 1000 1 2.641
## 703 0030992349 3.279 1997 1000 2 2.924
```

```

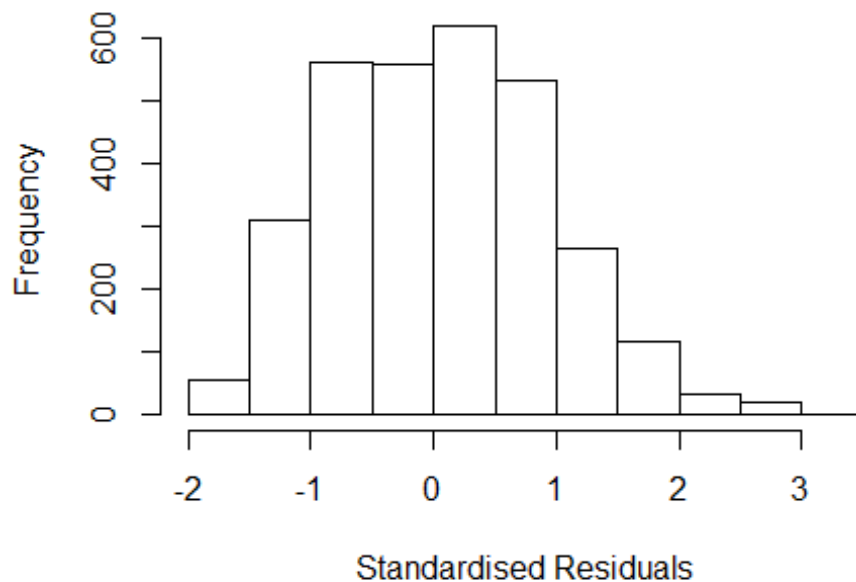
## 705 0030989227 2.636 1997 1000 2 2.513
## 752 0030934532 3.116 1997 1000 1 2.761
## 753 0030939391 3.026 1997 1000 2 2.671
## 803 0031035644 2.963 1997 1000 2 2.608
## 1030 0032572722 3.822 1998 1000 2 2.949
## 1035 0032552114 3.760 1998 1000 2 2.887
## 1080 0032474853 3.416 1998 1000 2 2.543
## 1102 0032511040 4.051 1998 1000 2 3.178
## 1407 0033607504 3.751 1999 1000 2 2.765
## 1434 0033615491 3.300 1999 1000 1 2.671
## 1473 0033539183 3.596 1999 1000 2 2.610
## 1534 0033542448 3.153 1999 1000 2 2.524
## 1556 0033587146 3.526 1999 1000 1 2.897
## 1800 0034676457 3.392 2000 1000 2 2.716
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8801 -0.6282  0.0394  0.6613  3.2059
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9585     0.1127   8.51 < 2e-16 ***
## FirstAuthorFemale1 0.3053     0.0402   7.59 4.2e-14 ***
## Year1997        -0.6460     0.1213  -5.32 1.1e-07 ***
## Year1998        -0.1134     0.1661  -0.68 0.49500
## Year1999        -0.3751     0.1527  -2.46 0.01408 *
## Year2000        -0.0823     0.1452  -0.57 0.57057
## Year2001        -0.1124     0.1637  -0.69 0.49240
## Year2002         0.4096     0.1217   3.37 0.00077 ***
## Year2003         0.3547     0.1349   2.63 0.00858 **
## Year2004         0.5230     0.1277   4.09 4.3e-05 ***
## Year2005         0.4168     0.1299   3.21 0.00135 **
## Year2006         0.4771     0.1285   3.71 0.00021 ***
## Year2007         0.5128     0.1314   3.90 9.7e-05 ***
## Year2008         0.5036     0.1360   3.70 0.00022 ***
## Year2009         0.5121     0.1350   3.79 0.00015 ***
## Year2010         0.4369     0.1411   3.10 0.00198 **
## Year2011         0.4840     0.1487   3.26 0.00114 **
## Year2012         0.6162     0.1372   4.49 7.3e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.669
## Multiple R-squared:  0.195, Adjusted R-squared:  0.19
## Convergence in 20 IRWLS iterations

```



```
##
## Robustness weights:
## observation 584 is an outlier with |weight| = 0 ( < 3.3e-05);
## 146 weights are ~= 1. The remaining 2927 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0095 0.7780 0.9020 0.8460 0.9770 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.25e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.031 1      1.015
## Year      1.031 16      1.001
```

### Residuals from last author



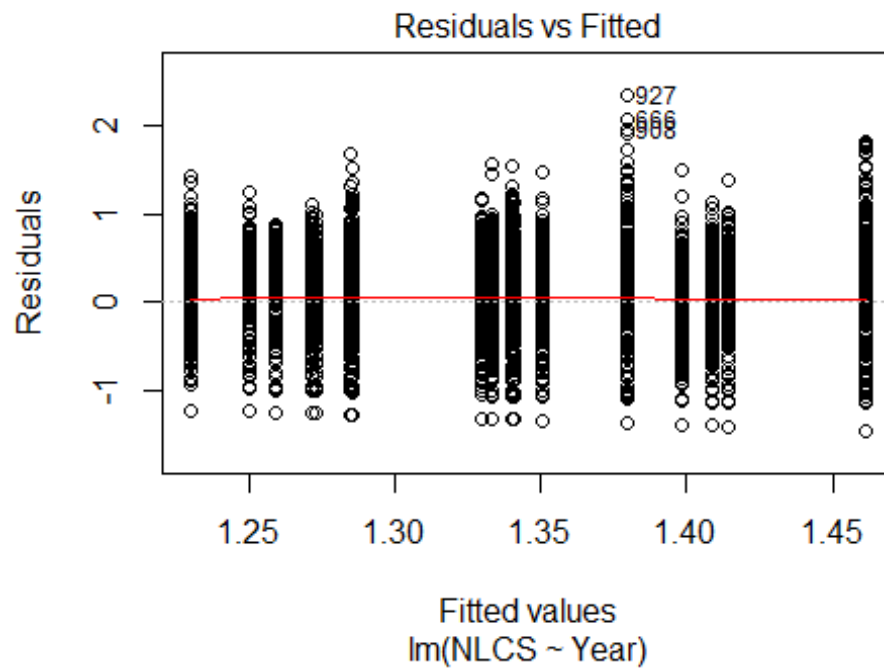
```
## [1] "List of 19 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 604  0030669546 3.039 1997    1000     2    2.684
## 664  0030795741 3.225 1997    1000     2    2.870
## 684  0030986236 2.880 1997    1000     1    2.525
## 691  0030878110 2.996 1997    1000     1    2.641
## 703  0030992349 3.279 1997    1000     2    2.924
## 705  0030989227 2.636 1997    1000     2    2.513
## 752  0030934532 3.116 1997    1000     1    2.761
## 753  0030939391 3.026 1997    1000     2    2.671
## 803  0031035644 2.963 1997    1000     2    2.608
## 1030 0032572722 3.822 1998    1000     2    2.949
## 1035 0032552114 3.760 1998    1000     2    2.887
## 1080 0032474853 3.416 1998    1000     2    2.543
## 1102 0032511040 4.051 1998    1000     2    3.178
## 1407 0033607504 3.751 1999    1000     2    2.765
## 1434 0033615491 3.300 1999    1000     1    2.671
## 1473 0033539183 3.596 1999    1000     2    2.610
## 1534 0033542448 3.153 1999    1000     2    2.524
## 1556 0033587146 3.526 1999    1000     1    2.897
## 1800 0034676457 3.392 2000    1000     2    2.716
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6790 -0.6877  0.0681  0.6667  3.1284
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0237     0.1103   9.28 < 2e-16 ***
## LastAuthorFemale1 -0.0956     0.0486  -1.96  0.04953 *
## Year1997          -0.6112     0.1209  -5.06  4.5e-07 ***
## Year1998          -0.1010     0.1608  -0.63  0.52976
## Year1999          -0.3360     0.1534  -2.19  0.02854 *
## Year2000          -0.0484     0.1425  -0.34  0.73412
## Year2001          -0.0861     0.1582  -0.54  0.58624
## Year2002           0.4374     0.1197   3.65  0.00026 ***
## Year2003           0.3898     0.1324   2.94  0.00327 **
## Year2004           0.5343     0.1252   4.27  2.0e-05 ***
## Year2005           0.4409     0.1271   3.47  0.00053 ***
## Year2006           0.5064     0.1253   4.04  5.5e-05 ***
## Year2007           0.5509     0.1291   4.27  2.0e-05 ***
## Year2008           0.5430     0.1327   4.09  4.4e-05 ***
## Year2009           0.5631     0.1318   4.27  2.0e-05 ***
## Year2010           0.4779     0.1386   3.45  0.00057 ***
## Year2011           0.5266     0.1470   3.58  0.00035 ***
```

```

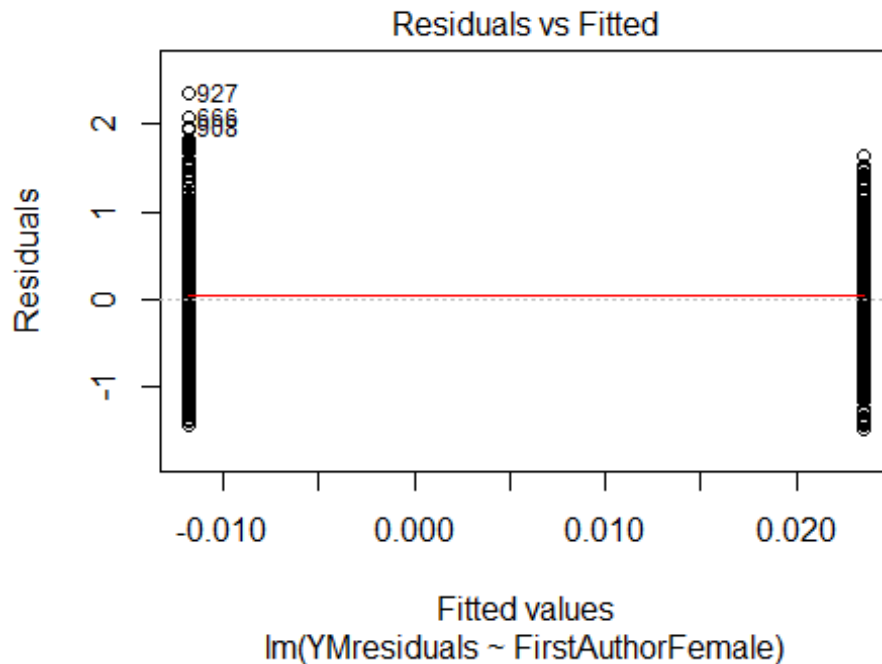
## Year2012          0.6553      0.1348      4.86  1.2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.168, Adjusted R-squared:  0.163
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 176 weights are ~= 1. The remaining 2898 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.008  0.807  0.908  0.854  0.969  0.999
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.25e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##           0          1000          0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3074"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1100"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   534  480  512  483  445  389  390  355  360  338  399  487  482  577  692
## 2011 2012
##   937 1312
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   241  216  239  273  200  118  274  229  243  230  272  341  349  440  550
## 2011 2012
##   755 1074
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   227  203  229  252  190  105  248  212  225  208  240  302  313  402  503
## 2011 2012

```

```
## 682 947
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 420, df = 16, p-value <2e-16
```

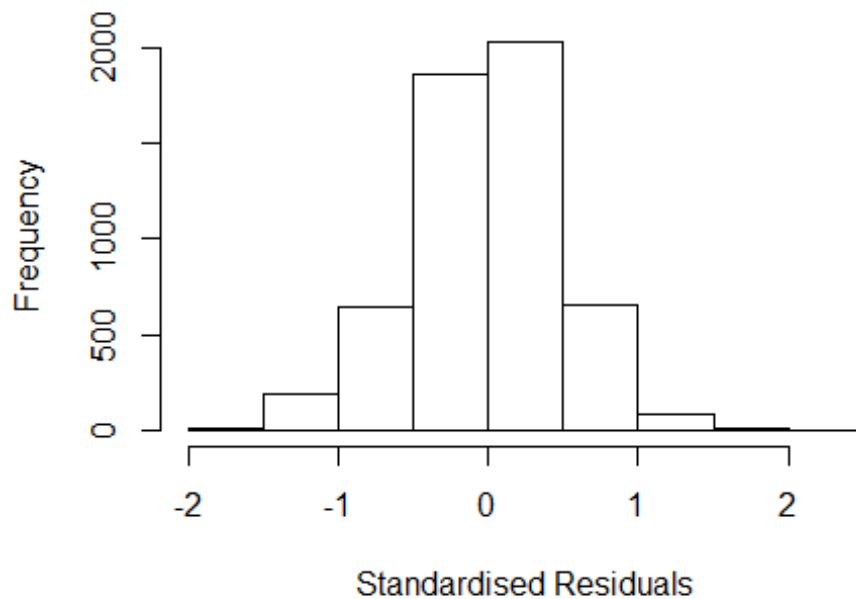


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.11625748029051"
## [1] "Male first author team size 2018 geometric mean: 3.40753317009802"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 63000, p-value = 3e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.6956079129643"
## [1] "Male last author team size 2018 geometric mean: 3.71943416469912"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.052 1          1.026
## LastAuthorFemale  1.037 1          1.018
## UniqueAuthors     1.241 4          1.027
## Year               1.273 16         1.008
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.73512 -0.29722  0.00702  0.30179  2.32418
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.26890    0.06912   18.36 < 2e-16 ***
## FirstAuthorFemale1 -0.01453    0.01390   -1.05  0.29587
## LastAuthorFemale1 -0.00818    0.01626   -0.50  0.61499
## UniqueAuthors2     0.35469    0.02608   13.60 < 2e-16 ***
## UniqueAuthors3     0.44977    0.02624   17.14 < 2e-16 ***
## UniqueAuthors4     0.47440    0.02770   17.12 < 2e-16 ***
## UniqueAuthors5     0.54100    0.02555   21.18 < 2e-16 ***
## Year1997         -0.14008    0.08925   -1.57  0.11661
## Year1998         -0.10870    0.08421   -1.29  0.19682
## Year1999         -0.25378    0.07783   -3.26  0.00112 **
```

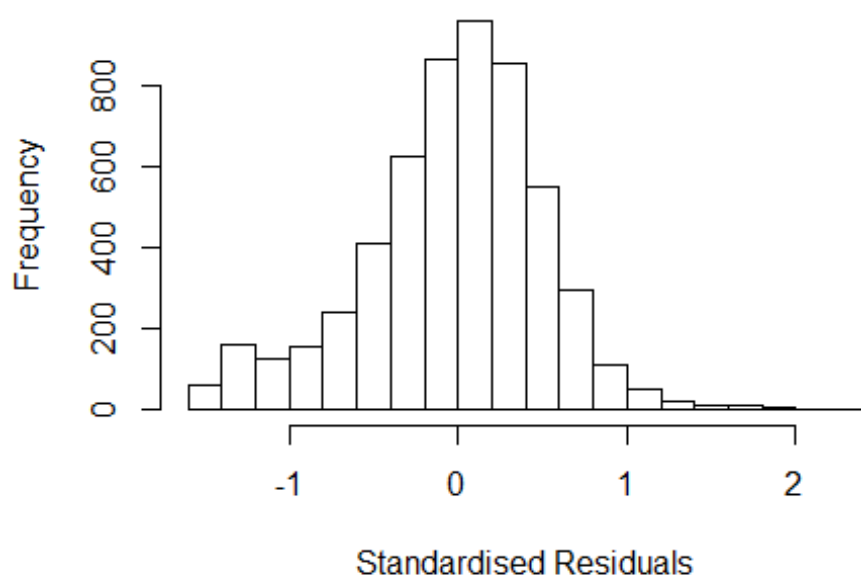
```

## Year2000      -0.25962      0.07863      -3.30      0.00097 ***
## Year2001      -0.28493      0.09006      -3.16      0.00157 **
## Year2002      -0.15779      0.07370      -2.14      0.03233 *
## Year2003      -0.24025      0.07455      -3.22      0.00128 **
## Year2004      -0.28623      0.07325      -3.91      9.4e-05 ***
## Year2005      -0.27773      0.07448      -3.73      0.00019 ***
## Year2006      -0.23938      0.07237      -3.31      0.00095 ***
## Year2007      -0.25211      0.07158      -3.52      0.00043 ***
## Year2008      -0.28117      0.07270      -3.87      0.00011 ***
## Year2009      -0.28644      0.07133      -4.02      6.0e-05 ***
## Year2010      -0.35836      0.07058      -5.08      4.0e-07 ***
## Year2011      -0.38786      0.06947      -5.58      2.5e-08 ***
## Year2012      -0.44922      0.06916      -6.50      9.0e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.451
## Multiple R-squared:  0.143, Adjusted R-squared:  0.14
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## observation 255 is an outlier with |weight| = 0 ( < 1.8e-05);
## 514 weights are ~ = 1. The remaining 4973 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0034 0.8550 0.9500 0.8900 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.82e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1 1.016
## LastAuthorFemale 1.033 1 1.016
## Year 1.063 16 1.002

```



## Residuals from first and last author



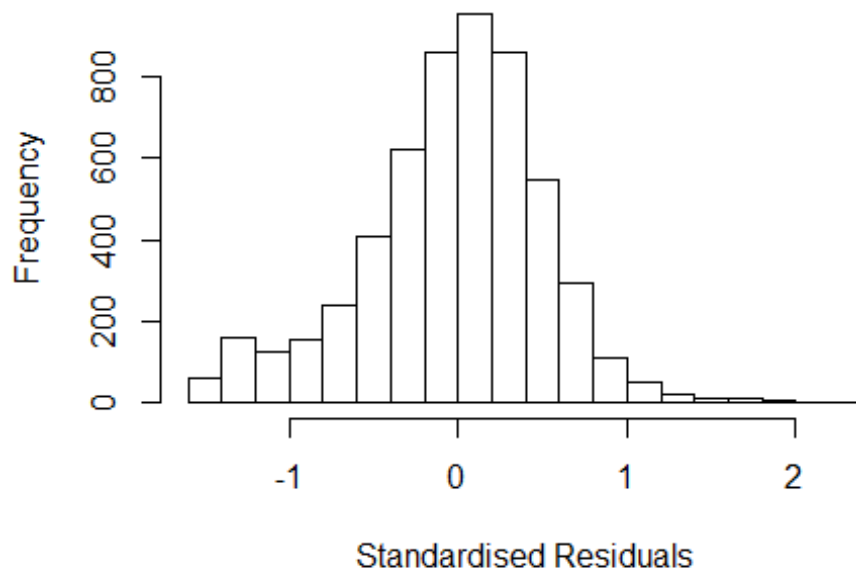
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4936 -0.3171 0.0246 0.3113 2.3760
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45282 0.06891 21.08 <2e-16 ***
## FirstAuthorFemale1 0.04341 0.01441 3.01 0.0026 **
## LastAuthorFemale1 -0.00264 0.01736 -0.15 0.8791
## Year1997 -0.10679 0.09503 -1.12 0.2611
## Year1998 -0.03726 0.08738 -0.43 0.6698
## Year1999 -0.14546 0.07970 -1.83 0.0680 .
## Year2000 -0.13532 0.08210 -1.65 0.0994 .
## Year2001 -0.14229 0.09414 -1.51 0.1307
## Year2002 -0.01074 0.07452 -0.14 0.8854
## Year2003 -0.08742 0.07641 -1.14 0.2526
## Year2004 -0.13830 0.07474 -1.85 0.0643 .
## Year2005 -0.09865 0.07575 -1.30 0.1929
```

```

## Year2006          -0.06425      0.07401      -0.87      0.3854
## Year2007          -0.05168      0.07218      -0.72      0.4740
## Year2008          -0.09274      0.07405      -1.25      0.2105
## Year2009          -0.10545      0.07246      -1.46      0.1457
## Year2010          -0.16948      0.07155      -2.37      0.0179 *
## Year2011          -0.17016      0.07005      -2.43      0.0152 *
## Year2012          -0.22413      0.06954      -3.22      0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.0144
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## observation 381 is an outlier with |weight| = 0 ( < 1.8e-05);
## 446 weights are ~= 1. The remaining 5041 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0043 0.8570 0.9510 0.8850 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.82e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1          1.016
## Year              1.033 16          1.001

```

## Residuals from first author



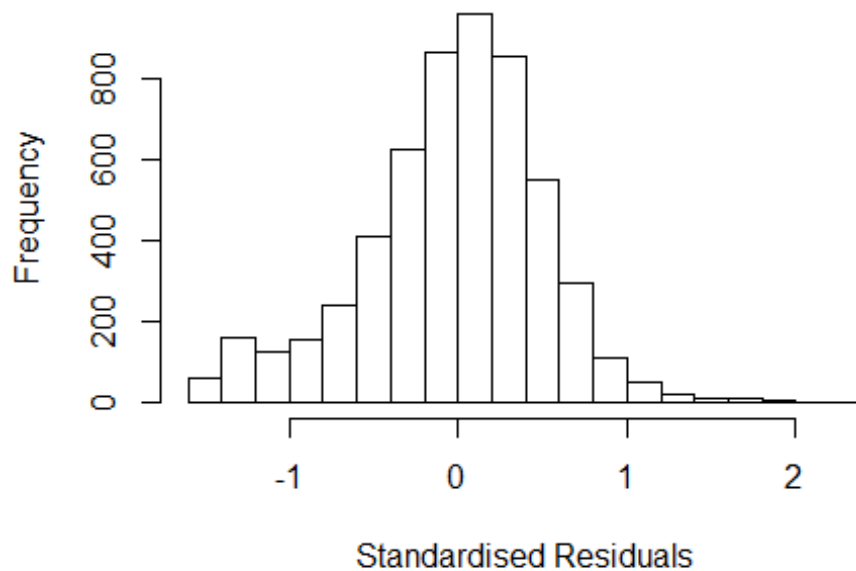
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4954 -0.3168 0.0248 0.3113 2.3764
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4523 0.0686 21.16 <2e-16 ***
## FirstAuthorFemale1 0.0431 0.0145 2.97 0.0030 **
## Year1997 -0.1067 0.0950 -1.12 0.2613
## Year1998 -0.0371 0.0872 -0.42 0.6709
## Year1999 -0.1453 0.0796 -1.83 0.0680 .
## Year2000 -0.1354 0.0821 -1.65 0.0991 .
## Year2001 -0.1421 0.0940 -1.51 0.1307
## Year2002 -0.0105 0.0744 -0.14 0.8880
## Year2003 -0.0872 0.0763 -1.14 0.2532
## Year2004 -0.1381 0.0746 -1.85 0.0642 .
## Year2005 -0.0985 0.0757 -1.30 0.1931
## Year2006 -0.0640 0.0739 -0.87 0.3865
```

```

## Year2007          -0.0516      0.0721   -0.72   0.4744
## Year2008          -0.0926      0.0740   -1.25   0.2109
## Year2009          -0.1055      0.0725   -1.46   0.1454
## Year2010          -0.1694      0.0715   -2.37   0.0179 *
## Year2011          -0.1701      0.0700   -2.43   0.0151 *
## Year2012          -0.2242      0.0695   -3.22   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.0146
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## observation 381 is an outlier with |weight| = 0 ( < 1.8e-05);
## 444 weights are ~= 1. The remaining 5043 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0044 0.8580 0.9510 0.8850 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

## Residuals from last author



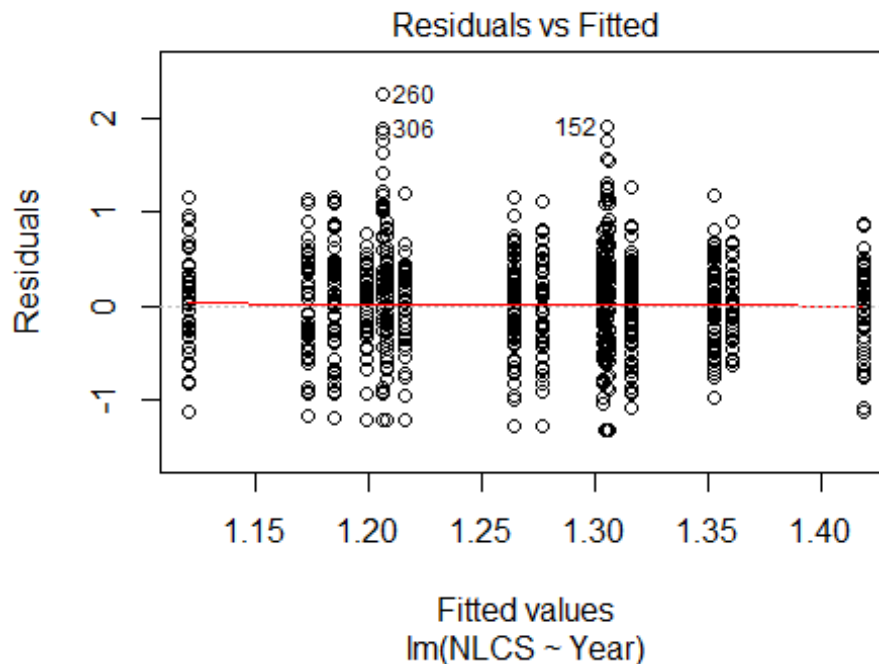
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4696 -0.3119 0.0219 0.3093 2.3687
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.46514 0.06834 21.44 <2e-16 ***
## LastAuthorFemale1 0.00446 0.01738 0.26 0.7976
## Year1997 -0.11183 0.09490 -1.18 0.2387
## Year1998 -0.04210 0.08745 -0.48 0.6302
## Year1999 -0.14819 0.07960 -1.86 0.0627 .
## Year2000 -0.13692 0.08202 -1.67 0.0951 .
## Year2001 -0.13999 0.09451 -1.48 0.1386
## Year2002 -0.01156 0.07433 -0.16 0.8764
## Year2003 -0.08941 0.07616 -1.17 0.2405
## Year2004 -0.14063 0.07456 -1.89 0.0593 .
## Year2005 -0.10010 0.07566 -1.32 0.1859
## Year2006 -0.06429 0.07381 -0.87 0.3838
```

```

## Year2007          -0.05343      0.07198      -0.74      0.4579
## Year2008          -0.09014      0.07387      -1.22      0.2224
## Year2009          -0.10422      0.07232      -1.44      0.1496
## Year2010          -0.16680      0.07143      -2.34      0.0196 *
## Year2011          -0.16605      0.06992      -2.37      0.0176 *
## Year2012          -0.21967      0.06943      -3.16      0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0162, Adjusted R-squared:  0.0132
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## observation 381 is an outlier with |weight| = 0 ( < 1.8e-05);
## 446 weights are ~= 1. The remaining 5041 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0052 0.8590 0.9500 0.8840 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.82e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5488"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1101"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 139 164 119 104 101 94 110 102 90 123 122 108 86 58 68
## 2011 2012
## 67 80
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 60 48 46 48 40 64 68 61 80 87 77 64 41 55

```

```
## 2011 2012
## 49 56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 57 45 46 47 36 55 63 55 71 80 68 56 35 52
## 2011 2012
## 44 50
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 180, df = 16, p-value <2e-16
```

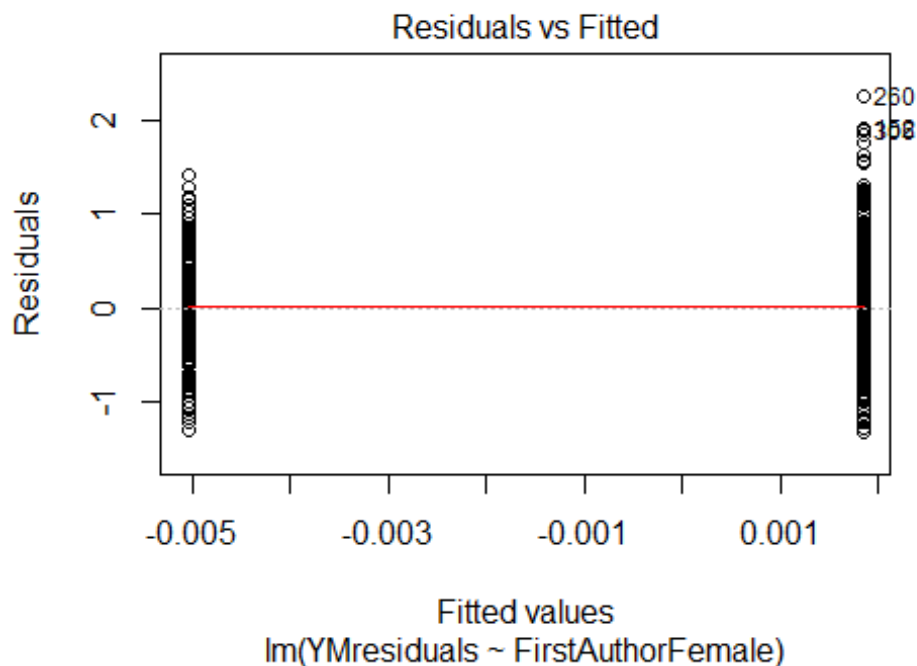


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.4, df = 1, p-value = 0.01
## [1] "Female first author team size 2018 geometric mean: 2.87089583095513"
## [1] "Male first author team size 2018 geometric mean: 2.27537237056012"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



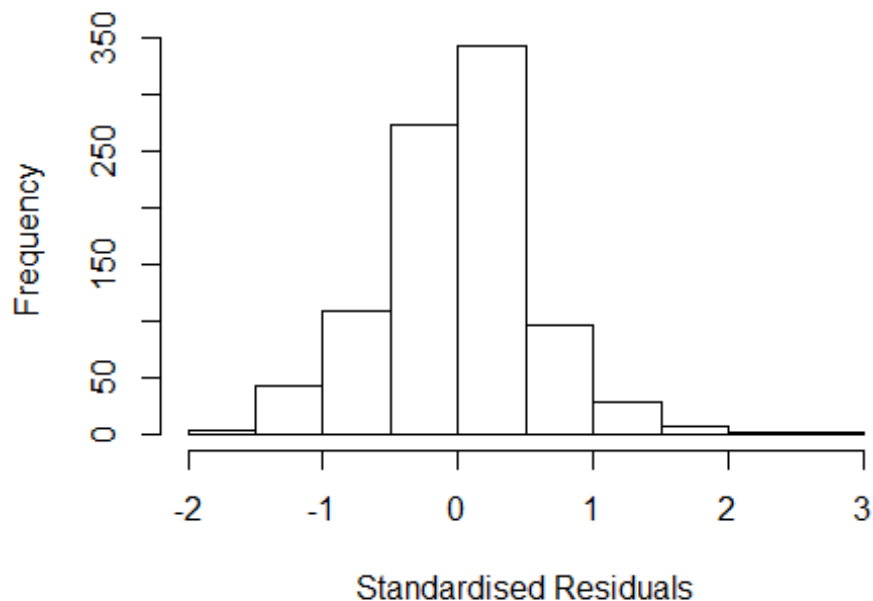
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.0530391109398"
## [1] "Male last author team size 2018 geometric mean: 2.54928920417129"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.269 1          1.127
## LastAuthorFemale  1.156 1          1.075
## UniqueAuthors    1.516 4          1.053
## Year              1.646 16         1.016
```

## Residuals from first and last author and team size



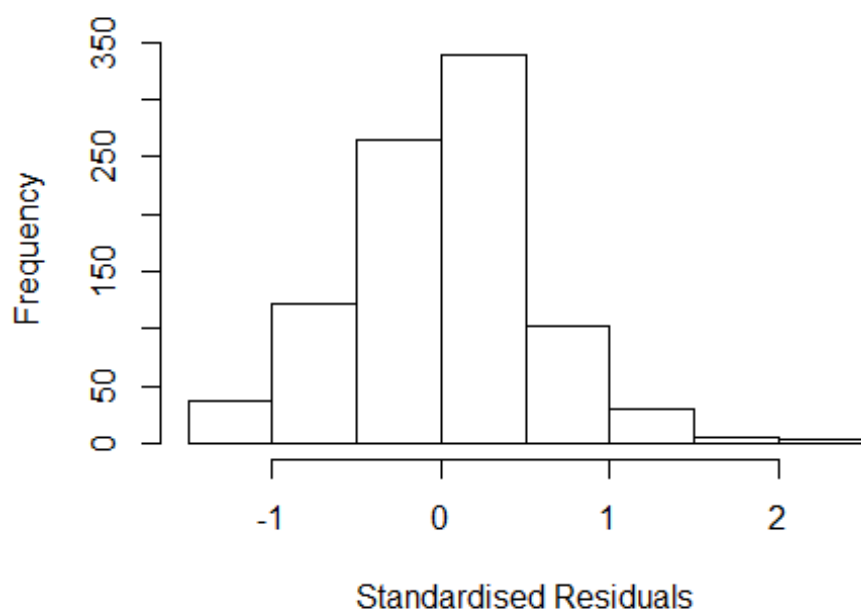
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 260 0031466983 3.453 1997    1100      5      2.58
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5008 -0.3223  0.0286  0.3171  2.5802
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22645    0.18083   6.78 2.2e-11 ***
## FirstAuthorFemale1 -0.04529    0.04082  -1.11   0.27
## LastAuthorFemale1  0.00177    0.04836   0.04   0.97
## UniqueAuthors2    0.23520    0.05090   4.62 4.4e-06 ***
## UniqueAuthors3    0.23591    0.05689   4.15 3.7e-05 ***
## UniqueAuthors4    0.25769    0.06237   4.13 3.9e-05 ***
## UniqueAuthors5    0.36879    0.06295   5.86 6.6e-09 ***
## Year1997        -0.35368    0.29161  -1.21   0.23
## Year1998         0.03843    0.24074   0.16   0.87
## Year1999        -0.21900    0.20072  -1.09   0.28
```

```

## Year2000      -0.21130      0.19644      -1.08      0.28
## Year2001      -0.16804      0.19944      -0.84      0.40
## Year2002      -0.19899      0.19724      -1.01      0.31
## Year2003      -0.16670      0.18971      -0.88      0.38
## Year2004      -0.17456      0.18733      -0.93      0.35
## Year2005      -0.16966      0.19053      -0.89      0.37
## Year2006      -0.12421      0.18736      -0.66      0.51
## Year2007      -0.05827      0.18575      -0.31      0.75
## Year2008      -0.09774      0.18969      -0.52      0.61
## Year2009      -0.14271      0.19743      -0.72      0.47
## Year2010      -0.15881      0.19223      -0.83      0.41
## Year2011      -0.07629      0.18836      -0.41      0.69
## Year2012       0.01406      0.19068      0.07      0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0834, Adjusted R-squared:  0.0605
## Convergence in 27 IRWLS iterations
##
## Robustness weights:
## observation 70 is an outlier with |weight| = 0 ( < 0.00011);
## 92 weights are ~= 1. The remaining 810 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8460 0.9510 0.8820 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.11e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.149 1 1.072
## LastAuthorFemale 1.111 1 1.054
## Year 1.151 16 1.004

```

## Residuals from first and last author

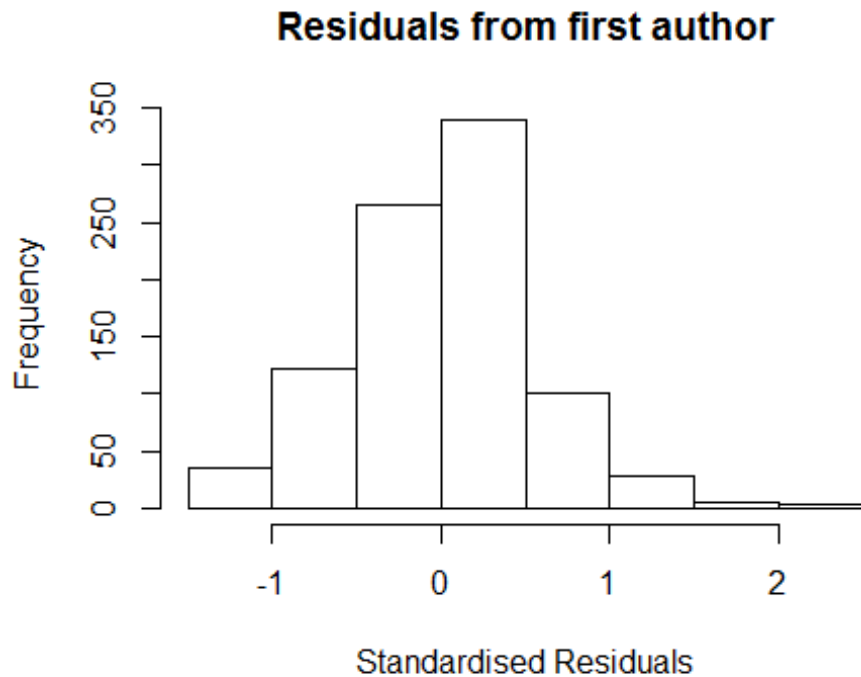


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3883 -0.3323 0.0352 0.3348 2.4735
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33748 0.19013 7.03 4e-12 ***
## FirstAuthorFemale1 0.01207 0.04034 0.30 0.76
## LastAuthorFemale1 0.01475 0.04998 0.30 0.77
## Year1997 -0.35795 0.30540 -1.17 0.24
## Year1998 0.02402 0.25165 0.10 0.92
## Year1999 -0.19405 0.21022 -0.92 0.36
## Year2000 -0.17574 0.20599 -0.85 0.39
## Year2001 -0.14324 0.20768 -0.69 0.49
## Year2002 -0.14251 0.20660 -0.69 0.49
## Year2003 -0.12261 0.19935 -0.62 0.54
## Year2004 -0.11120 0.19573 -0.57 0.57
## Year2005 -0.08654 0.19704 -0.44 0.66
```

```

## Year2006      -0.03567    0.19532   -0.18    0.86
## Year2007      0.02443    0.19501    0.13    0.90
## Year2008     -0.02954    0.19657   -0.15    0.88
## Year2009     -0.06020    0.20436   -0.29    0.77
## Year2010     -0.07172    0.20075   -0.36    0.72
## Year2011     -0.00273    0.19856   -0.01    0.99
## Year2012      0.08970    0.20011    0.45    0.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0341, Adjusted R-squared:  0.0144
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## observation 70 is an outlier with |weight| = 0 ( < 0.00011);
## 78 weights are ~= 1. The remaining 824 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.026  0.857   0.950   0.886   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.11e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.088 1          1.043
## Year              1.088 16          1.003

```

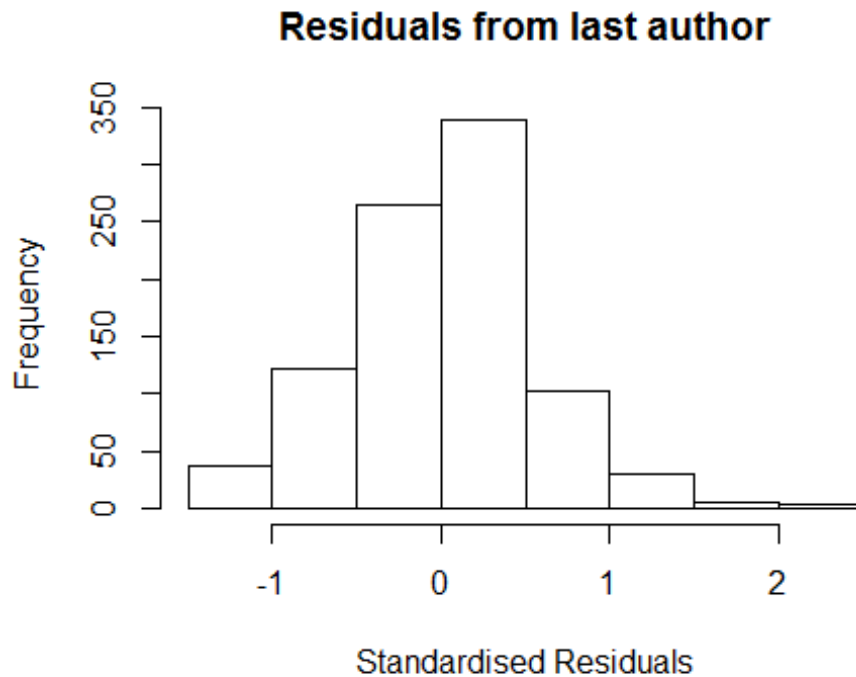


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.380 -0.331 0.035 0.341 2.473
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.33981 0.18949 7.07 3.1e-12 ***
## FirstAuthorFemale1 0.01536 0.03928 0.39 0.70
## Year1997 -0.35998 0.30600 -1.18 0.24
## Year1998 0.02468 0.25274 0.10 0.92
## Year1999 -0.19532 0.21005 -0.93 0.35
## Year2000 -0.17529 0.20637 -0.85 0.40
## Year2001 -0.14443 0.20751 -0.70 0.49
## Year2002 -0.14401 0.20629 -0.70 0.49
## Year2003 -0.12335 0.19942 -0.62 0.54
## Year2004 -0.11219 0.19567 -0.57 0.57
## Year2005 -0.08713 0.19722 -0.44 0.66
## Year2006 -0.03710 0.19506 -0.19 0.85
```

```

## Year2007          0.02412      0.19526      0.12      0.90
## Year2008          -0.03090      0.19646     -0.16      0.88
## Year2009          -0.06124      0.20404     -0.30      0.76
## Year2010          -0.07261      0.20072     -0.36      0.72
## Year2011          -0.00412      0.19835     -0.02      0.98
## Year2012           0.08990      0.20040      0.45      0.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0342, Adjusted R-squared:  0.0156
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## observation 70 is an outlier with |weight| = 0 ( < 0.00011);
## 78 weights are ~= 1. The remaining 824 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0254 0.8570 0.9490 0.8860 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.11e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.051 1          1.025
## Year              1.051 16          1.002

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.383 -0.336  0.035  0.340  2.472
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.339869   0.189427   7.07 3.1e-12 ***
## LastAuthorFemale1 0.018679   0.048620   0.38  0.70
## Year1997       -0.358720   0.305659  -1.17  0.24
## Year1998        0.024027   0.251097   0.10  0.92
## Year1999       -0.193536   0.209792  -0.92  0.36
## Year2000       -0.176651   0.205485  -0.86  0.39
## Year2001       -0.143069   0.207225  -0.69  0.49
## Year2002       -0.142622   0.206114  -0.69  0.49
## Year2003       -0.123338   0.198835  -0.62  0.54
## Year2004       -0.111384   0.195150  -0.57  0.57
## Year2005       -0.086204   0.196484  -0.44  0.66
## Year2006       -0.034906   0.194715  -0.18  0.86
```

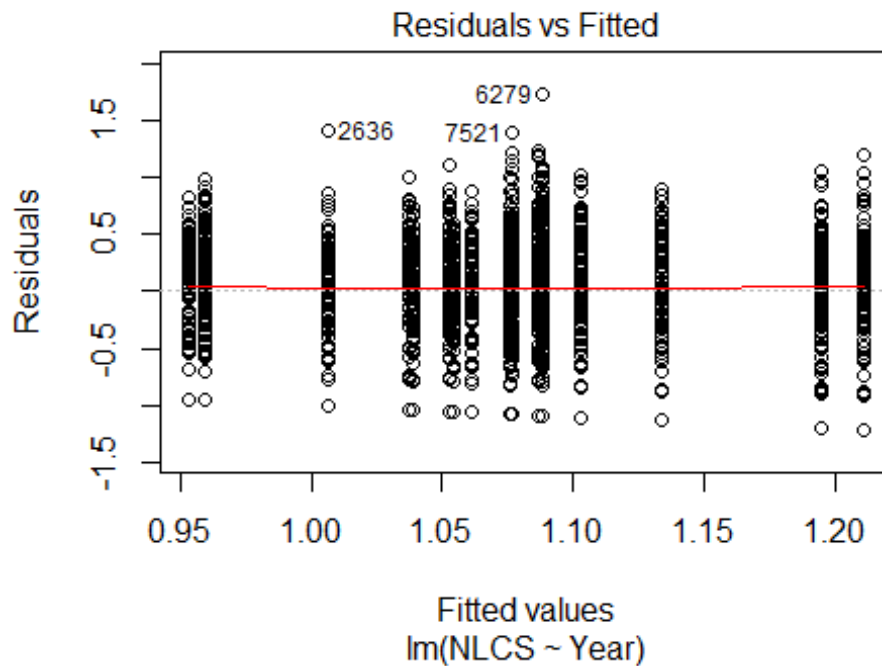


```

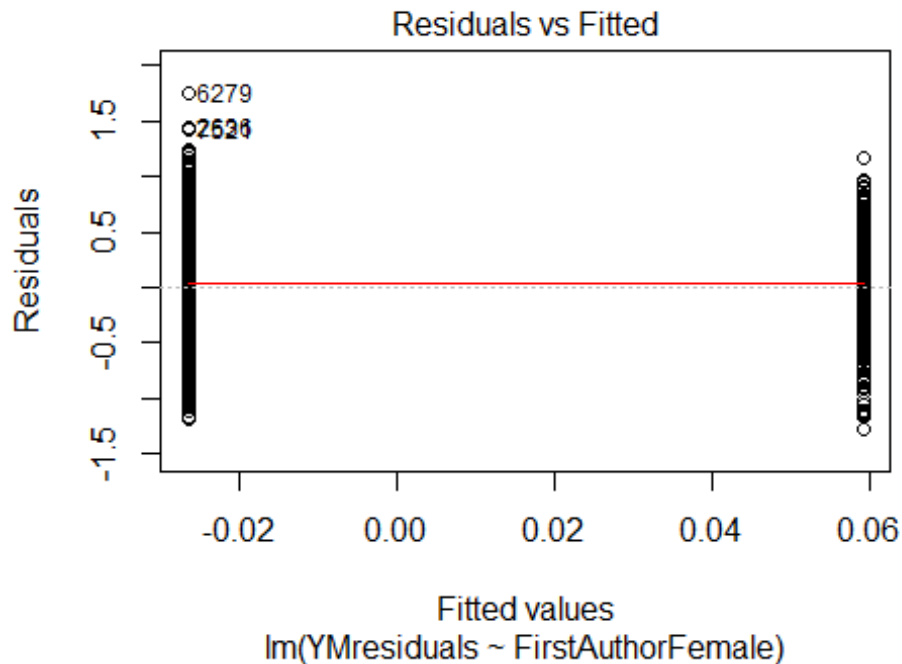
## Year2007      0.024826   0.194375   0.13   0.90
## Year2008     -0.027892   0.195843  -0.14   0.89
## Year2009     -0.059149   0.203640  -0.29   0.77
## Year2010     -0.070429   0.199929  -0.35   0.72
## Year2011     -0.000871   0.197680   0.00   1.00
## Year2012      0.089526   0.199540   0.45   0.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.0341, Adjusted R-squared:  0.0155
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## observation 70 is an outlier with |weight| = 0 ( < 0.00011);
## 77 weights are ~= 1. The remaining 825 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0256 0.8570 0.9490 0.8860 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.11e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 903"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1102"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 559 500 479 493 467 400 433 326 298 268 236 276 260 226 243
## 2011 2012
## 258 233
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 131 102 92 131 114 105 176 100 91 100 92 112 110 111 128

```

```
## 2011 2012
## 122 125
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 120 93 81 120 106 93 141 87 76 89 85 103 101 103 119
## 2011 2012
## 109 117
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 63, df = 16, p-value = 2e-07
```

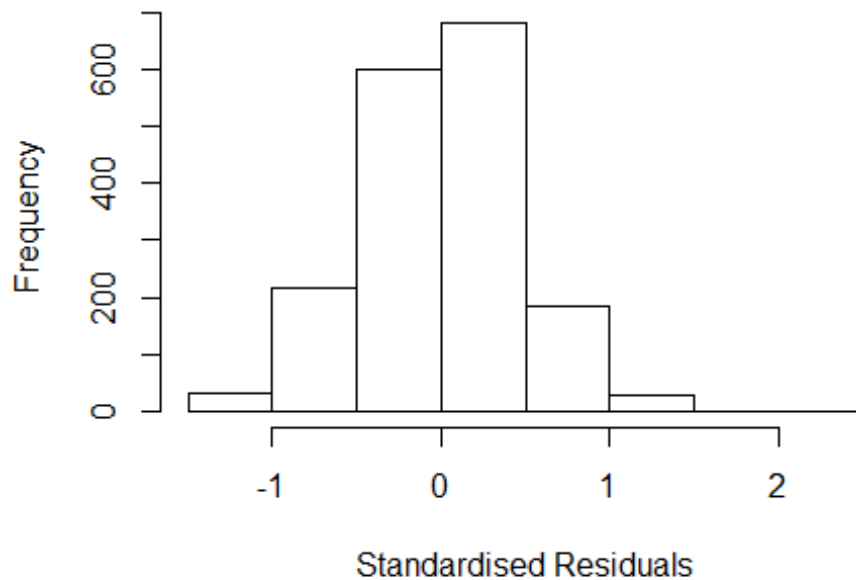


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 26, df = 1, p-value = 3e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 3.88329640687872"
## [1] "Male first author team size 2018 geometric mean: 3.25796547162851"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.57193702300231"
## [1] "Male last author team size 2018 geometric mean: 3.48186806371034"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.086 1          1.042
## LastAuthorFemale  1.077 1          1.038
## UniqueAuthors    1.317 4          1.035
## Year              1.349 16         1.009
```

## Residuals from first and last author and team size



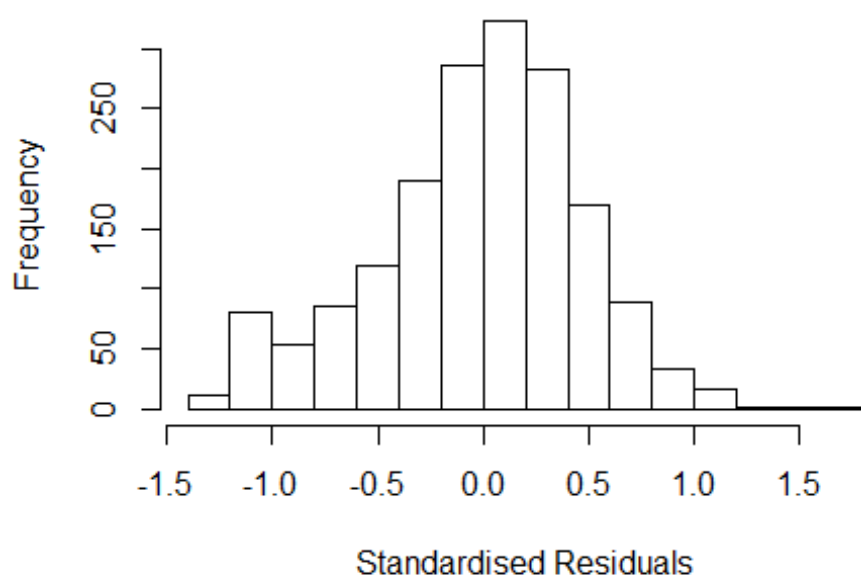
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3671 -0.2893 0.0166 0.3004 2.0278
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.81922 0.05946 13.78 < 2e-16 ***
## FirstAuthorFemale1 0.05266 0.02331 2.26 0.0240 *
## LastAuthorFemale1 0.02603 0.03220 0.81 0.4189
## UniqueAuthors2 0.21484 0.04521 4.75 2.2e-06 ***
## UniqueAuthors3 0.35494 0.04317 8.22 3.9e-16 ***
## UniqueAuthors4 0.41910 0.04413 9.50 < 2e-16 ***
## UniqueAuthors5 0.49524 0.04387 11.29 < 2e-16 ***
## Year1997 0.00155 0.07520 0.02 0.9836
## Year1998 0.04839 0.06800 0.71 0.4769
## Year1999 -0.00947 0.06095 -0.16 0.8766
```

```

## Year2000      -0.11366    0.06167   -1.84    0.0655 .
## Year2001      -0.09047    0.06390   -1.42    0.1570
## Year2002      -0.16716    0.05934   -2.82    0.0049 **
## Year2003      -0.10963    0.05829   -1.88    0.0602 .
## Year2004      -0.09375    0.06484   -1.45    0.1484
## Year2005      -0.19708    0.06843   -2.88    0.0040 **
## Year2006      -0.07577    0.06849   -1.11    0.2687
## Year2007      -0.07006    0.06977   -1.00    0.3154
## Year2008      -0.03598    0.07069   -0.51    0.6109
## Year2009       0.07130    0.06523    1.09    0.2745
## Year2010       0.04189    0.06170    0.68    0.4973
## Year2011      -0.03277    0.07251   -0.45    0.6514
## Year2012      -0.10227    0.06992   -1.46    0.1438
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.156, Adjusted R-squared:  0.145
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1605 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0011 0.8630 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.74e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061 1 1.030
## LastAuthorFemale 1.070 1 1.035
## Year 1.134 16 1.004

```

## Residuals from first and last author



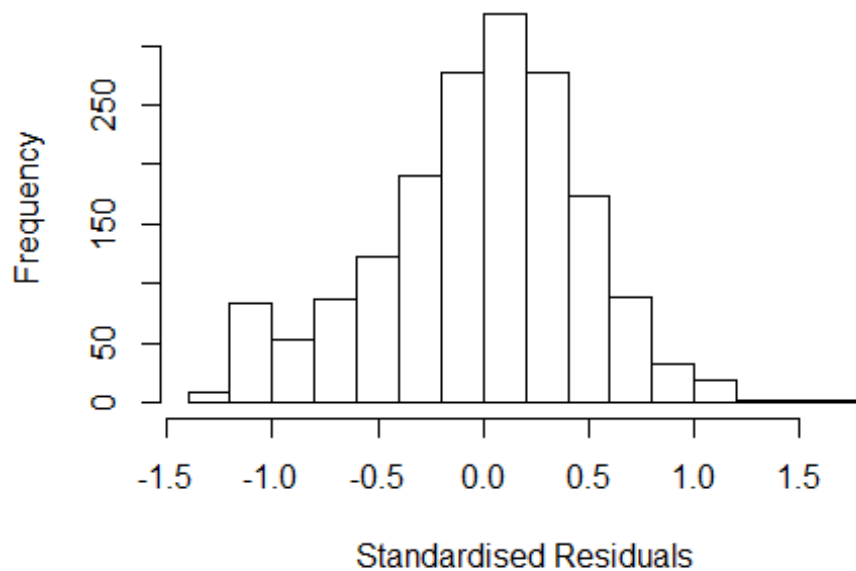
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3344 -0.3073 0.0262 0.3118 1.7064
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.090462 0.047542 22.94 < 2e-16 ***
## FirstAuthorFemale1 0.100490 0.023725 4.24 2.4e-05 ***
## LastAuthorFemale1 0.019152 0.031729 0.60 0.5462
## Year1997 -0.003174 0.074897 -0.04 0.9662
## Year1998 0.050481 0.070812 0.71 0.4760
## Year1999 -0.022907 0.061225 -0.37 0.7083
## Year2000 -0.109821 0.060193 -1.82 0.0683 .
## Year2001 -0.072766 0.064687 -1.12 0.2608
## Year2002 -0.166580 0.064151 -2.60 0.0095 **
## Year2003 -0.050111 0.059234 -0.85 0.3977
## Year2004 -0.048966 0.065684 -0.75 0.4561
## Year2005 -0.158702 0.070778 -2.24 0.0251 *
```

```

## Year2006      -0.049500    0.071976   -0.69    0.4917
## Year2007      -0.042829    0.072563   -0.59    0.5551
## Year2008       0.014165    0.075806    0.19    0.8518
## Year2009       0.124335    0.067974    1.83    0.0675 .
## Year2010       0.092354    0.064388    1.43    0.1517
## Year2011       0.000492    0.076050    0.01    0.9948
## Year2012      -0.044879    0.073509   -0.61    0.5416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.0367, Adjusted R-squared:  0.0266
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.119  0.857   0.950   0.893   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1      1.029
## Year              1.059 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3188 -0.3057 0.0268 0.3133 1.7036
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09208 0.04747 23.00 < 2e-16 ***
## FirstAuthorFemale1 0.10219 0.02399 4.26 2.2e-05 ***
## Year1997 -0.00277 0.07486 -0.04 0.9705
## Year1998 0.05414 0.06989 0.77 0.4387
## Year1999 -0.02202 0.06129 -0.36 0.7194
## Year2000 -0.10888 0.06025 -1.81 0.0709 .
## Year2001 -0.07199 0.06474 -1.11 0.2663
## Year2002 -0.16583 0.06418 -2.58 0.0099 **
## Year2003 -0.05005 0.05931 -0.84 0.3988
## Year2004 -0.04864 0.06567 -0.74 0.4590
## Year2005 -0.15812 0.07084 -2.23 0.0257 *
## Year2006 -0.04874 0.07186 -0.68 0.4977
```

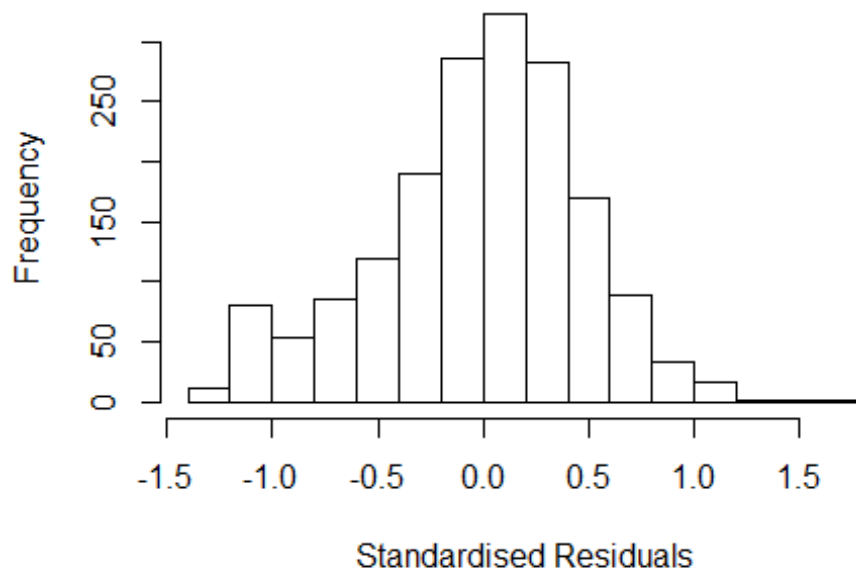


```

## Year2007          -0.04285    0.07253   -0.59    0.5547
## Year2008          0.01537    0.07576    0.20    0.8393
## Year2009          0.12455    0.06792    1.83    0.0668 .
## Year2010          0.09324    0.06439    1.45    0.1478
## Year2011          0.00261    0.07575    0.03    0.9725
## Year2012         -0.04324    0.07330   -0.59    0.5553
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.0363, Adjusted R-squared:  0.0268
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.857   0.950   0.893   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.068 1          1.033
## Year              1.068 16          1.002

```

## Residuals from last author



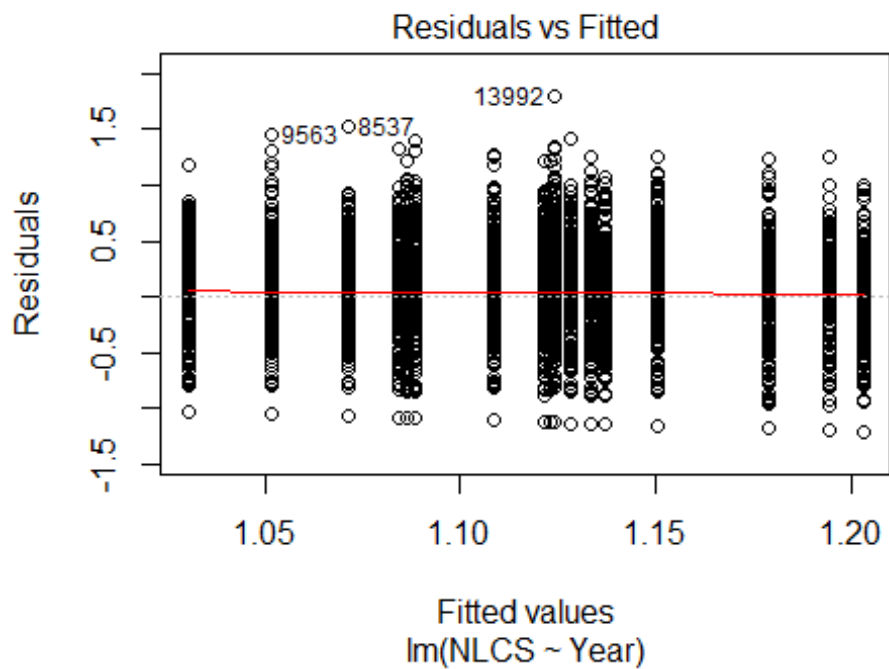
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2886 -0.3026 0.0262 0.3115 1.6737
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12e+00 4.67e-02 24.00 <2e-16 ***
## LastAuthorFemale1 3.37e-02 3.22e-02 1.05 0.295
## Year1997 -1.16e-02 7.40e-02 -0.16 0.875
## Year1998 5.25e-02 7.09e-02 0.74 0.459
## Year1999 -2.80e-02 6.13e-02 -0.46 0.648
## Year2000 -1.18e-01 6.02e-02 -1.96 0.050 .
## Year2001 -6.97e-02 6.52e-02 -1.07 0.285
## Year2002 -1.65e-01 6.53e-02 -2.52 0.012 *
## Year2003 -4.77e-02 5.88e-02 -0.81 0.418
## Year2004 -4.32e-02 6.53e-02 -0.66 0.508
## Year2005 -1.64e-01 7.09e-02 -2.32 0.021 *
## Year2006 -4.55e-02 7.24e-02 -0.63 0.530
```

```

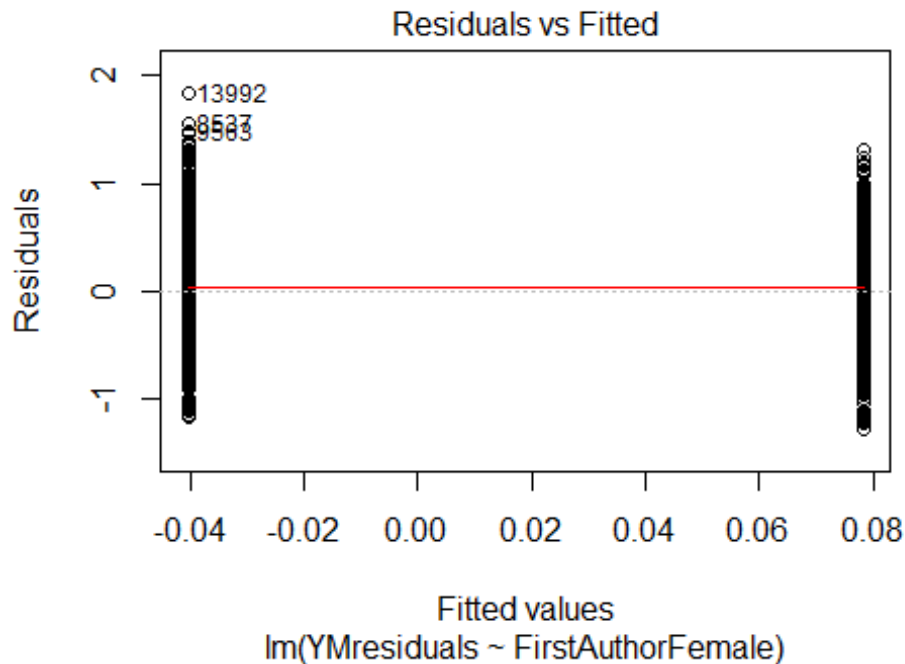
## Year2007          -4.16e-02   7.31e-02   -0.57   0.569
## Year2008          1.68e-02   7.63e-02    0.22   0.826
## Year2009          1.34e-01   6.78e-02    1.98   0.048 *
## Year2010          1.01e-01   6.48e-02    1.56   0.119
## Year2011          -6.27e-05   7.65e-02    0.00   0.999
## Year2012          -4.37e-02   7.31e-02   -0.60   0.550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.453
## Multiple R-squared:  0.0278, Adjusted R-squared:  0.0182
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1602 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.142  0.860  0.948  0.893  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1743"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1103"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  719  753  771  727  680  719  639  543  566  558  597  553  578  584  555
## 2011 2012
##  567  564
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  222  259  221  228  181  198  283  255  284  261  305  307  307  323  292
## 2011 2012

```

```
## 305 318
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 205 234 201 208 160 177 248 227 248 227 281 272 283 303 261
## 2011 2012
## 276 298
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 44, df = 16, p-value = 2e-04
```

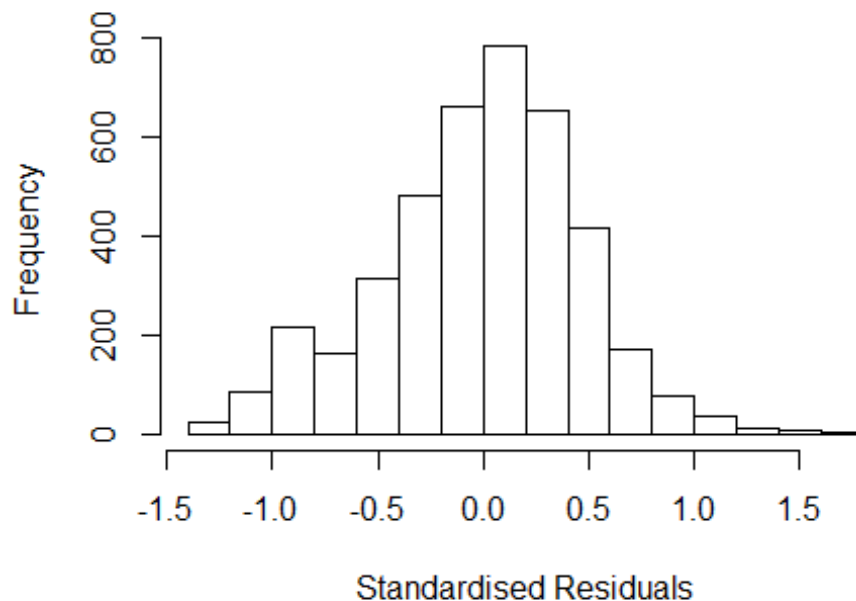


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 57, df = 1, p-value = 4e-14
```



```
## [1] "Female first author team size 2018 geometric mean: 2.95890047444995"
## [1] "Male first author team size 2018 geometric mean: 2.39751263569209"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.82973341819621"
## [1] "Male last author team size 2018 geometric mean: 2.56134109294894"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1          1.041
## LastAuthorFemale  1.046 1          1.023
## UniqueAuthors     1.183 4          1.021
## Year              1.201 16         1.006
```

## Residuals from first and last author and team size



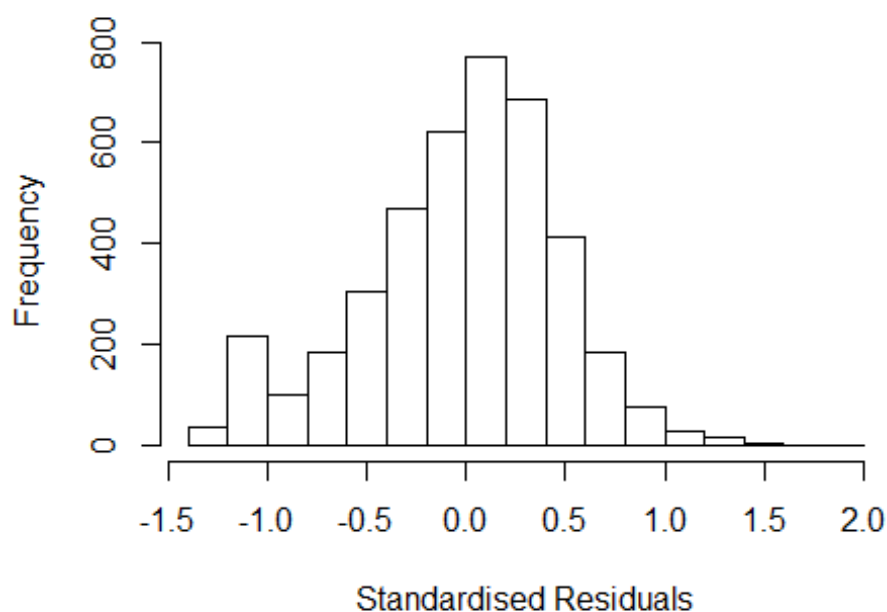
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3813 -0.3059 0.0242 0.3040 1.7117
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89477 0.03703 24.17 < 2e-16 ***
## FirstAuthorFemale1 0.07376 0.01523 4.84 1.3e-06 ***
## LastAuthorFemale1 0.06322 0.01775 3.56 0.00037 ***
## UniqueAuthors2 0.24995 0.02277 10.98 < 2e-16 ***
## UniqueAuthors3 0.29048 0.02303 12.62 < 2e-16 ***
## UniqueAuthors4 0.34425 0.02752 12.51 < 2e-16 ***
## UniqueAuthors5 0.39636 0.02701 14.67 < 2e-16 ***
## Year1997 0.05431 0.04960 1.09 0.27360
## Year1998 0.03802 0.04936 0.77 0.44116
## Year1999 0.07019 0.04736 1.48 0.13837
```

```

## Year2000          0.06690    0.05046    1.33  0.18493
## Year2001          0.06797    0.04546    1.50  0.13498
## Year2002         -0.00470    0.04327   -0.11  0.91351
## Year2003         -0.06255    0.04615   -1.36  0.17536
## Year2004         -0.05984    0.04484   -1.33  0.18207
## Year2005         -0.11994    0.04686   -2.56  0.01052 *
## Year2006         -0.05540    0.04478   -1.24  0.21614
## Year2007         -0.10752    0.04596   -2.34  0.01937 *
## Year2008         -0.02117    0.04341   -0.49  0.62588
## Year2009         -0.02581    0.04342   -0.59  0.55223
## Year2010          0.00426    0.04680    0.09  0.92744
## Year2011         -0.04686    0.04830   -0.97  0.33206
## Year2012         -0.06971    0.04567   -1.53  0.12700
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.105
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 337 weights are ~= 1. The remaining 3772 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.088  0.860  0.948  0.893  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.064 1      1.032
## LastAuthorFemale  1.042 1      1.021
## Year              1.058 16      1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3375 -0.3014 0.0348 0.3039 1.8525
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.050806 0.035889 29.28 < 2e-16 ***
## FirstAuthorFemale1 0.121994 0.015536 7.85 5.2e-15 ***
## LastAuthorFemale1 0.051548 0.017825 2.89 0.0038 **
## Year1997 0.051513 0.050502 1.02 0.3078
## Year1998 0.053011 0.049212 1.08 0.2815
## Year1999 0.099089 0.048273 2.05 0.0402 *
## Year2000 0.113133 0.051060 2.22 0.0268 *
## Year2001 0.107198 0.046317 2.31 0.0207 *
## Year2002 0.037109 0.044230 0.84 0.4015
## Year2003 0.000774 0.047211 0.02 0.9869
## Year2004 -0.000438 0.046243 -0.01 0.9924
## Year2005 -0.065403 0.049072 -1.33 0.1827
```

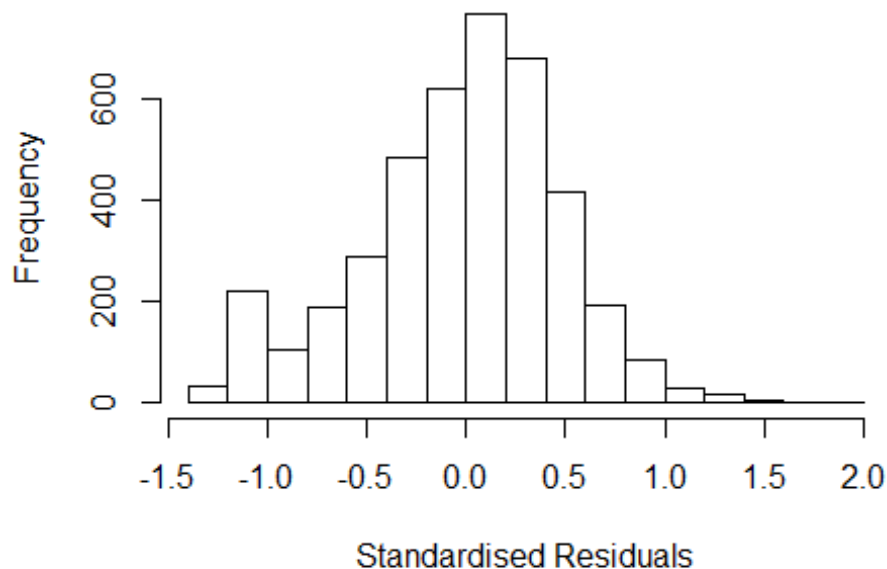


```

## Year2006          0.006835    0.045832    0.15    0.8815
## Year2007         -0.042953    0.047783   -0.90    0.3687
## Year2008          0.041108    0.044669    0.92    0.3575
## Year2009          0.042803    0.045025    0.95    0.3418
## Year2010          0.077728    0.048512    1.60    0.1092
## Year2011          0.034629    0.050672    0.68    0.4944
## Year2012          0.014719    0.046191    0.32    0.7500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.028, Adjusted R-squared:  0.0238
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 334 weights are ~= 1. The remaining 3775 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0423 0.8550 0.9500 0.8890 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```

## Residuals from first author

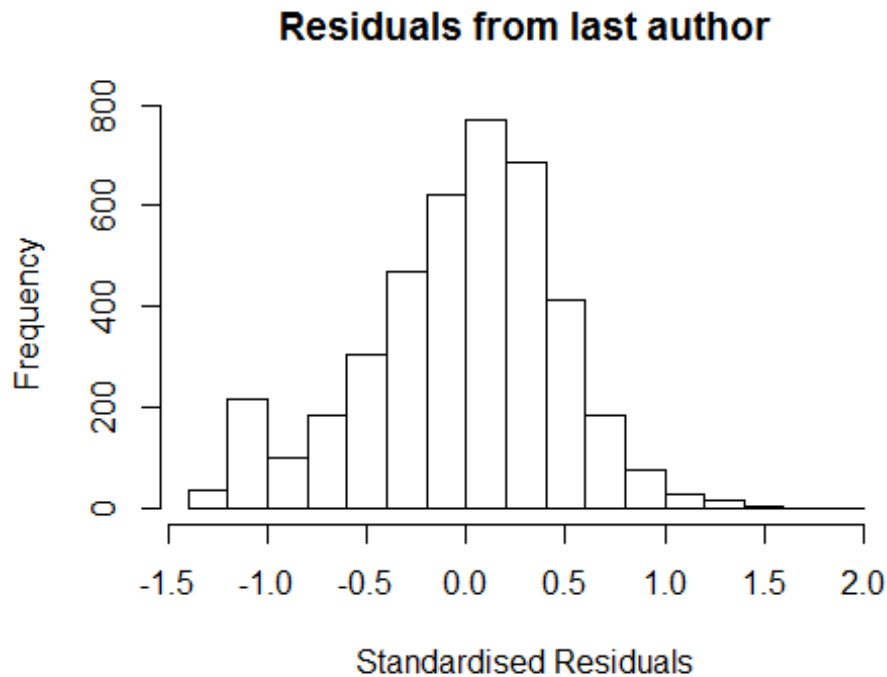


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3065 -0.3060 0.0357 0.3052 1.8431
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.059342 0.035583 29.77 <2e-16 ***
## FirstAuthorFemale1 0.134281 0.015502 8.66 <2e-16 ***
## Year1997 0.049483 0.050403 0.98 0.326
## Year1998 0.050612 0.048933 1.03 0.301
## Year1999 0.095088 0.048098 1.98 0.048 *
## Year2000 0.112926 0.051033 2.21 0.027 *
## Year2001 0.107995 0.046181 2.34 0.019 *
## Year2002 0.034797 0.044051 0.79 0.430
## Year2003 0.000355 0.047339 0.01 0.994
## Year2004 -0.003168 0.046162 -0.07 0.945
## Year2005 -0.066683 0.049068 -1.36 0.174
## Year2006 0.005963 0.045783 0.13 0.896
```

```

## Year2007      -0.044488    0.047650   -0.93    0.351
## Year2008      0.040371    0.044626    0.90    0.366
## Year2009      0.043416    0.045068    0.96    0.335
## Year2010      0.075282    0.048459    1.55    0.120
## Year2011      0.033931    0.050775    0.67    0.504
## Year2012      0.015512    0.046372    0.33    0.738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.0217
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 335 weights are ~= 1. The remaining 3774 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0459 0.8550 0.9500 0.8890 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.011
## Year      1.021 16      1.001

```



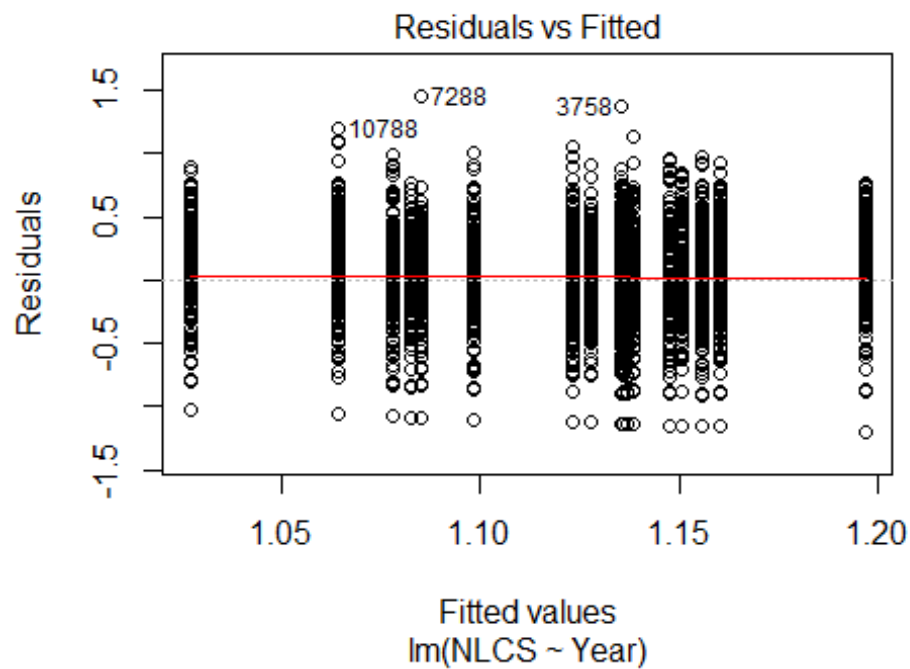
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2845 -0.3070  0.0419  0.3029  1.8058
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0759     0.0359  30.00 < 2e-16 ***
## LastAuthorFemale1  0.0888     0.0177   5.02 5.5e-07 ***
## Year1997          0.0542     0.0506   1.07  0.284
## Year1998          0.0534     0.0498   1.07  0.283
## Year1999          0.0979     0.0485   2.02  0.044 *
## Year2000          0.1199     0.0517   2.32  0.021 *
## Year2001          0.1121     0.0468   2.39  0.017 *
## Year2002          0.0485     0.0445   1.09  0.277
## Year2003          0.0128     0.0471   0.27  0.785
## Year2004          0.0101     0.0466   0.22  0.829
## Year2005         -0.0516     0.0498  -1.04  0.300
## Year2006          0.0100     0.0460   0.22  0.827
```

```

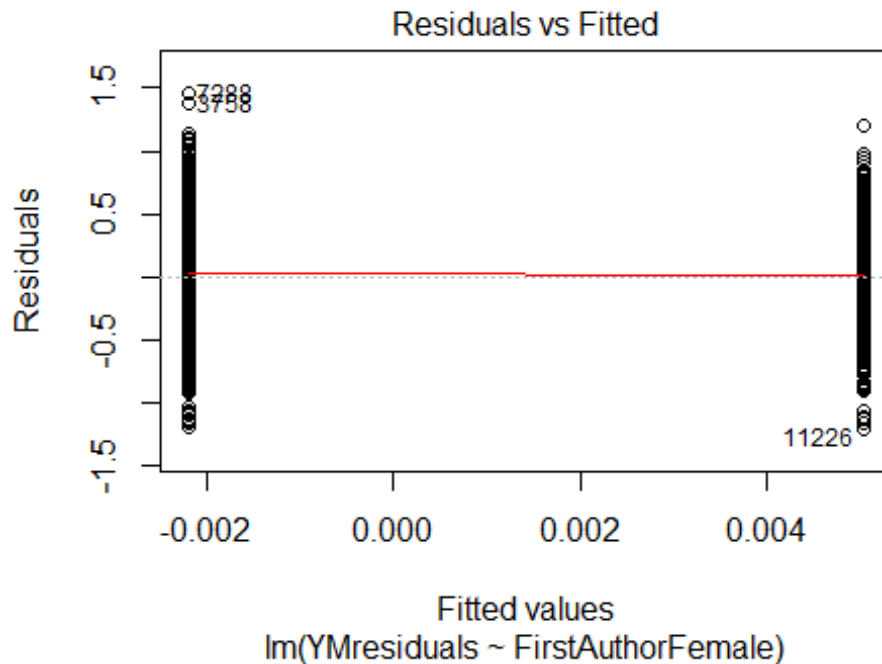
## Year2007          -0.0299      0.0480    -0.62      0.534
## Year2008           0.0623      0.0450      1.39      0.166
## Year2009           0.0583      0.0451      1.29      0.197
## Year2010           0.0986      0.0490      2.01      0.044 *
## Year2011           0.0439      0.0507      0.87      0.386
## Year2012           0.0364      0.0465      0.78      0.434
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.0105
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 305 weights are ~= 1. The remaining 3804 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0612 0.8560 0.9500 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4109"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1104"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 618 606 598 610 568 589 543 452 374 403 420 479 448 437 431
## 2011 2012
## 400 412
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 196 173 185 224 175 162 246 205 180 198 218 262 235 244 229
## 2011 2012

```

```
## 223 230
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 179 161 165 213 152 148 217 187 165 185 191 237 211 222 208
## 2011 2012
## 203 207
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```

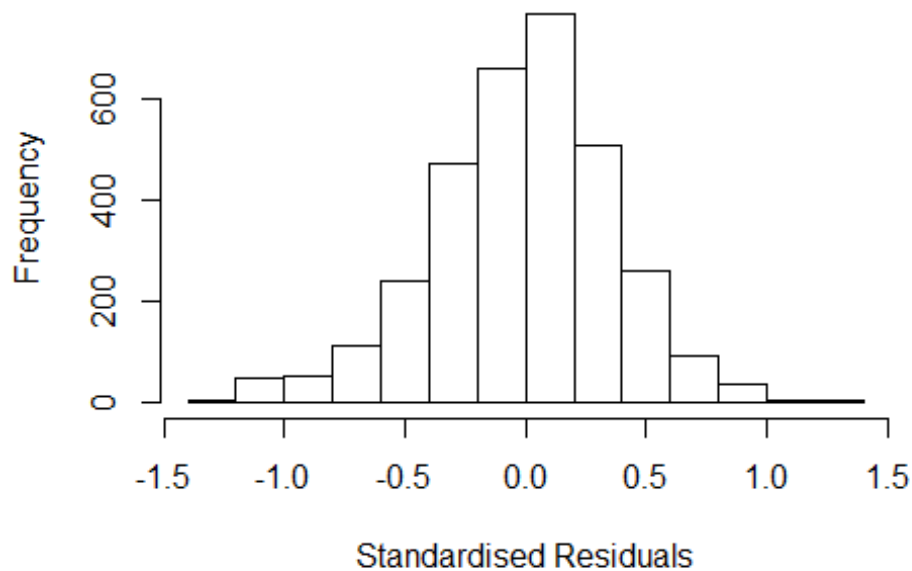


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 3.37862921133667"
## [1] "Male first author team size 2018 geometric mean: 3.36262571793839"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6100, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.42562938539472"
## [1] "Male last author team size 2018 geometric mean: 3.35143870281348"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4900, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1          1.021
## LastAuthorFemale  1.040 1          1.020
## UniqueAuthors    1.148 4          1.017
## Year              1.193 16         1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3137 -0.2380  0.0107  0.2282  1.3415
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00702    0.03283   30.67 < 2e-16 ***
## FirstAuthorFemale1 -0.00800    0.01409   -0.57  0.570
## LastAuthorFemale1 -0.03115    0.01600   -1.95  0.052 .
## UniqueAuthors2     0.12934    0.02332    5.55 3.1e-08 ***
## UniqueAuthors3     0.16464    0.02312    7.12 1.3e-12 ***
## UniqueAuthors4     0.19886    0.02502    7.95 2.6e-15 ***
## UniqueAuthors5     0.27756    0.02464   11.26 < 2e-16 ***
## Year1997          0.04313    0.04112    1.05  0.294
## Year1998         -0.00435    0.04180   -0.10  0.917
## Year1999          0.03317    0.03801    0.87  0.383
```

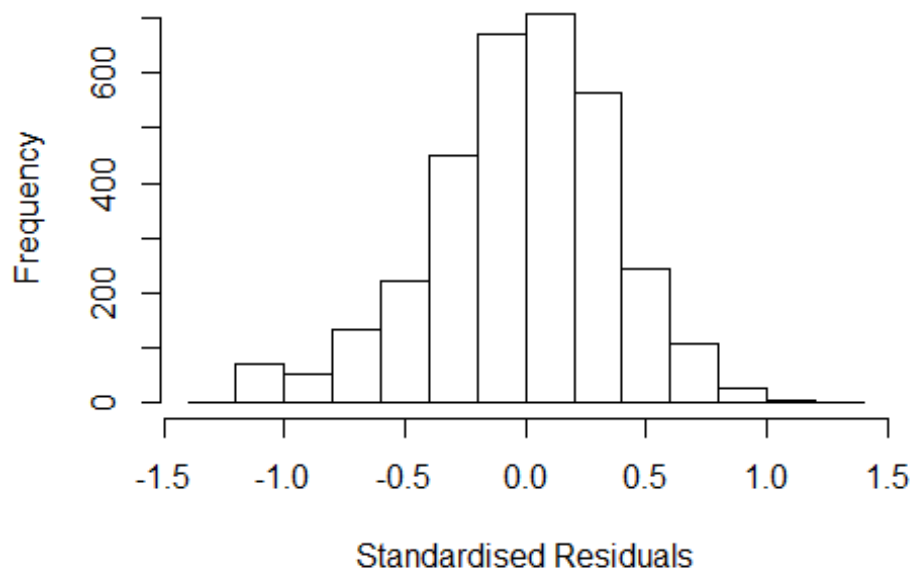


```

## Year2000          0.01479    0.04246    0.35    0.728
## Year2001          0.02912    0.04292    0.68    0.497
## Year2002          0.01085    0.03788    0.29    0.775
## Year2003         -0.07334    0.03982   -1.84    0.066 .
## Year2004         -0.03907    0.03743   -1.04    0.297
## Year2005         -0.00552    0.03655   -0.15    0.880
## Year2006         -0.06331    0.03710   -1.71    0.088 .
## Year2007          0.00596    0.03566    0.17    0.867
## Year2008         -0.01693    0.03715   -0.46    0.649
## Year2009         -0.00215    0.03633   -0.06    0.953
## Year2010         -0.00862    0.03789   -0.23    0.820
## Year2011         -0.10273    0.03987   -2.58    0.010 *
## Year2012          0.05191    0.03865    1.34    0.179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.0622, Adjusted R-squared:  0.0558
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 290 weights are ~= 1. The remaining 2961 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.864  0.950  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.08e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## LastAuthorFemale  1.035 1          1.017
## Year              1.049 16          1.002

```

## Residuals from first and last author



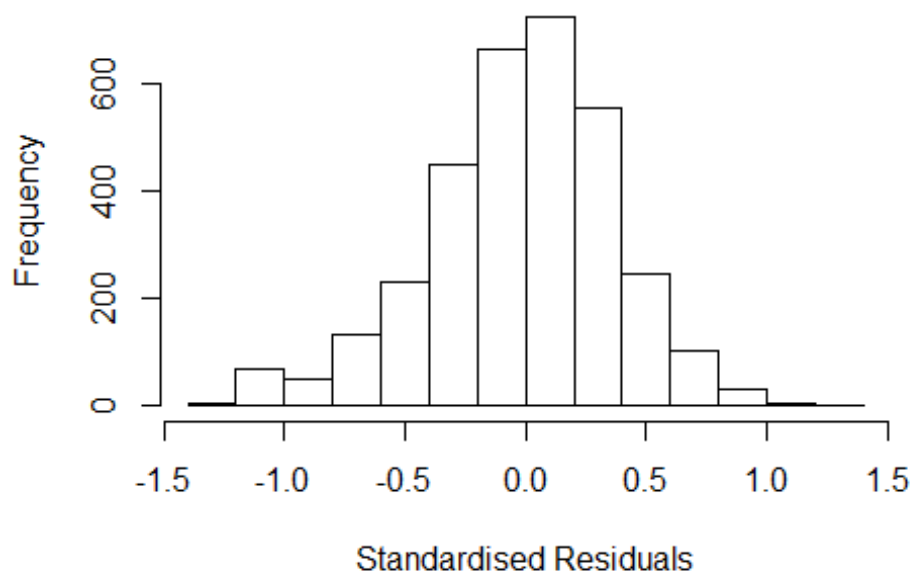
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22791 -0.24143  0.00717  0.24211  1.34424
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.12688    0.02899   38.87  <2e-16 ***
## FirstAuthorFemale1  0.00637    0.01427    0.45  0.6553
## LastAuthorFemale1 -0.04003    0.01620   -2.47  0.0135 *
## Year1997          0.04233    0.04214    1.00  0.3152
## Year1998         -0.01079    0.04343   -0.25  0.8038
## Year1999          0.03306    0.03903    0.85  0.3971
## Year2000          0.02959    0.04372    0.68  0.4986
## Year2001          0.03588    0.04463    0.80  0.4215
## Year2002          0.03739    0.03892    0.96  0.3369
## Year2003         -0.05197    0.04060   -1.28  0.2007
## Year2004         -0.00482    0.03919   -0.12  0.9022
## Year2005          0.02205    0.03794    0.58  0.5611
```

```

## Year2006      -0.03130    0.03793   -0.83    0.4093
## Year2007      0.05647    0.03650    1.55    0.1219
## Year2008      0.02258    0.03770    0.60    0.5492
## Year2009      0.04281    0.03772    1.13    0.2565
## Year2010      0.03942    0.03926    1.00    0.3155
## Year2011     -0.05525    0.04115   -1.34    0.1795
## Year2012      0.10103    0.03892    2.60    0.0095 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0133, Adjusted R-squared:  0.00779
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 282 weights are ~= 1. The remaining 2969 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.123  0.868  0.950  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1      1.013
## Year              1.026 16      1.001

```

## Residuals from first author



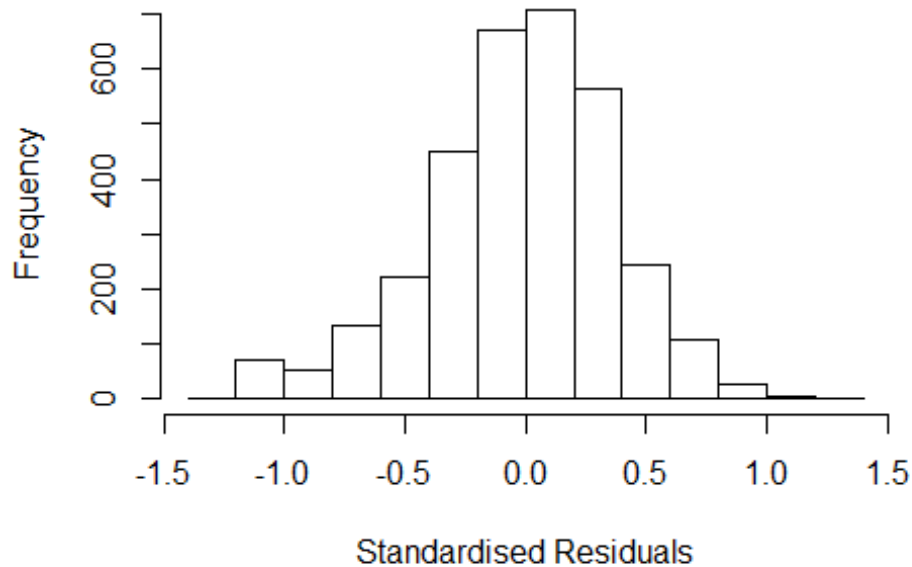
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2226 -0.2424  0.0082  0.2398  1.3500
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11895    0.02874   38.93  <2e-16 ***
## FirstAuthorFemale1 0.00107    0.01423    0.07  0.9403
## Year1997        0.04208    0.04227    1.00  0.3195
## Year1998       -0.01011    0.04351   -0.23  0.8163
## Year1999        0.03448    0.03895    0.89  0.3761
## Year2000        0.03096    0.04379    0.71  0.4796
## Year2001        0.03800    0.04464    0.85  0.3946
## Year2002        0.04085    0.03896    1.05  0.2945
## Year2003       -0.05325    0.04068   -1.31  0.1907
## Year2004       -0.00306    0.03933   -0.08  0.9380
## Year2005        0.02471    0.03781    0.65  0.5135
## Year2006       -0.03206    0.03796   -0.84  0.3984
```

```

## Year2007          0.05619    0.03649    1.54    0.1236
## Year2008          0.02240    0.03769    0.59    0.5523
## Year2009          0.04472    0.03758    1.19    0.2341
## Year2010          0.04097    0.03933    1.04    0.2976
## Year2011         -0.05238    0.04105   -1.28    0.2020
## Year2012          0.10263    0.03888    2.64    0.0083 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.0115, Adjusted R-squared:  0.00627
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 271 weights are ~= 1. The remaining 2980 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.868  0.951  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.012
## Year            1.023 16          1.001

```

## Residuals from last author



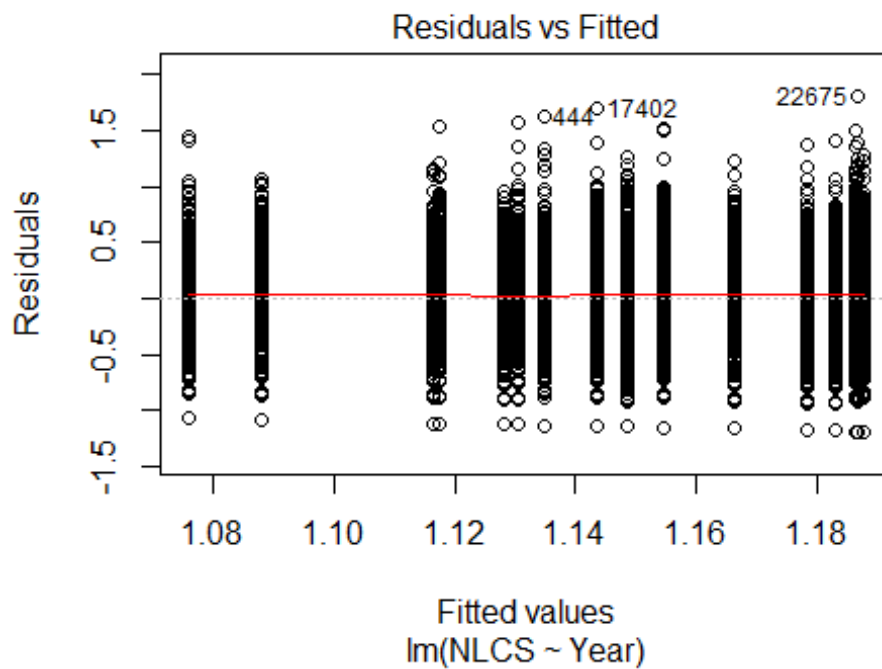
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23052 -0.24074  0.00751  0.24081  1.34277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12807    0.02884   39.12  <2e-16 ***
## LastAuthorFemale1 -0.03893    0.01613   -2.41  0.0158 *
## Year1997         0.04252    0.04217    1.01  0.3134
## Year1998        -0.00991    0.04330   -0.23  0.8190
## Year1999         0.03341    0.03899    0.86  0.3916
## Year2000         0.02999    0.04367    0.69  0.4923
## Year2001         0.03616    0.04458    0.81  0.4174
## Year2002         0.03751    0.03891    0.96  0.3352
## Year2003        -0.05158    0.04059   -1.27  0.2039
## Year2004        -0.00447    0.03917   -0.11  0.9092
## Year2005         0.02263    0.03793    0.60  0.5509
## Year2006        -0.03061    0.03787   -0.81  0.4189
```

```

## Year2007      0.05704      0.03649      1.56      0.1181
## Year2008      0.02312      0.03766      0.61      0.5394
## Year2009      0.04339      0.03771      1.15      0.2499
## Year2010      0.03982      0.03925      1.01      0.3104
## Year2011     -0.05413      0.04104     -1.32      0.1873
## Year2012      0.10244      0.03877      2.64      0.0083 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.356
## Multiple R-squared:  0.0132, Adjusted R-squared:  0.00803
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 273 weights are ~= 1. The remaining 2978 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.125  0.868  0.951  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3251"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1105"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1108 1041 1061 988 1076 1058 1000 880 825 873 880 956 1046 1173 1091
## 2011 2012
## 1125 1033
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 466 485 482 507 461 391 620 567 513 531 574 640 723 825 756
## 2011 2012

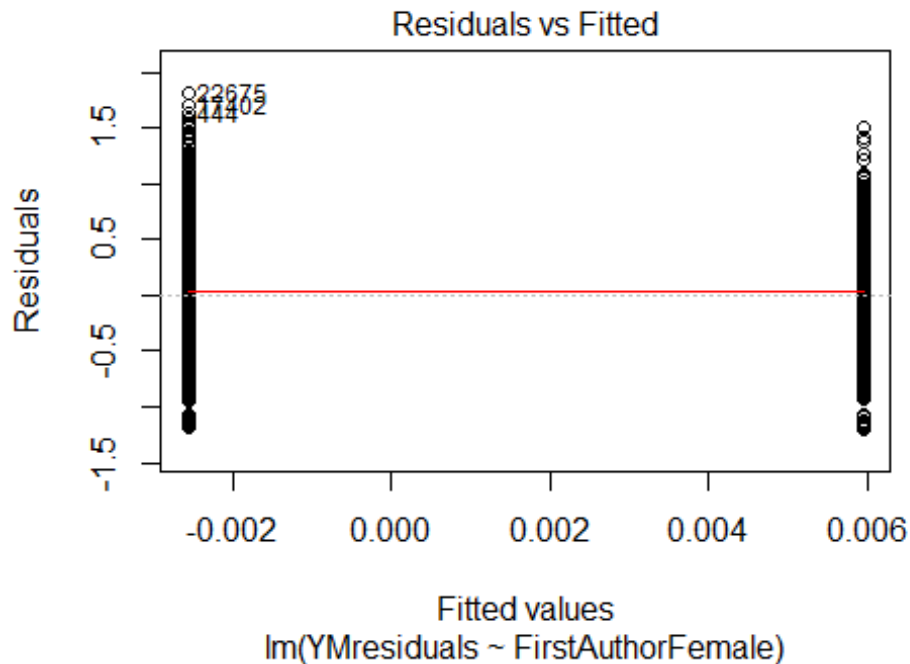
```

```
## 824 772
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 431 448 451 476 419 362 550 515 472 482 525 581 653 761 683
## 2011 2012
## 748 714
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 60, df = 16, p-value = 6e-07
```



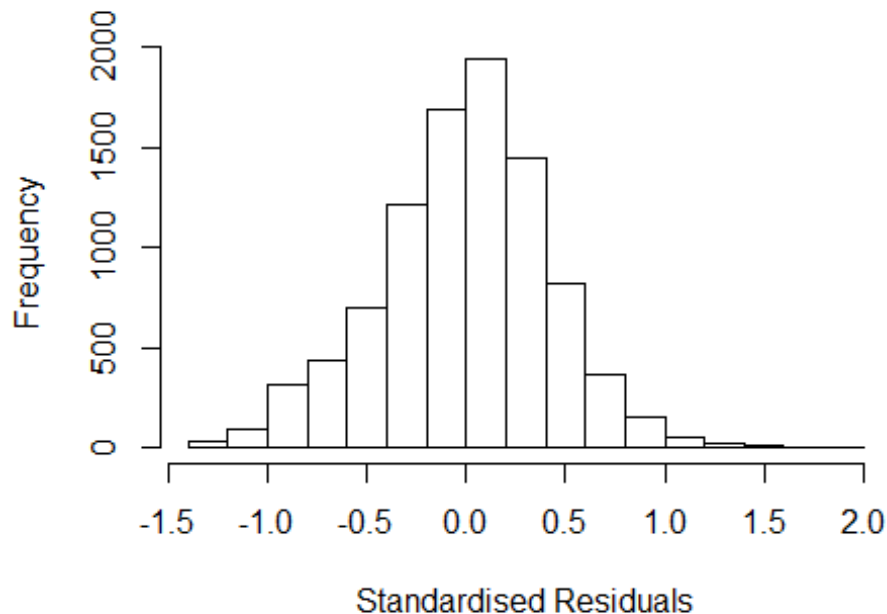
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 47, df = 1, p-value = 8e-12
```





```
## [1] "Female first author team size 2018 geometric mean: 3.3529078572026"
## [1] "Male first author team size 2018 geometric mean: 2.94181388807442"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 49000, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.14851258537318"
## [1] "Male last author team size 2018 geometric mean: 3.08130069781503"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 35000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## LastAuthorFemale  1.015 1          1.007
## UniqueAuthors    1.147 4          1.017
## Year             1.159 16          1.005
```

## Residuals from first and last author and team size



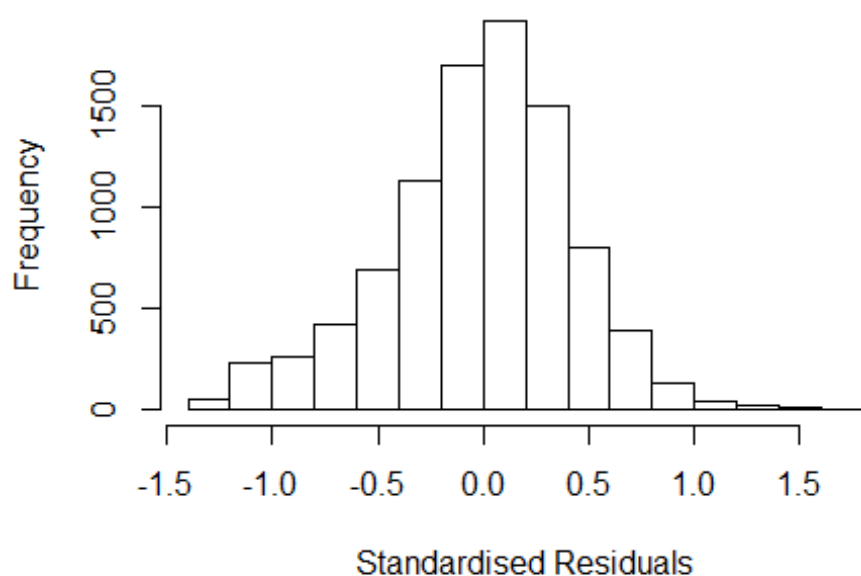
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3353 -0.2702  0.0162  0.2629  1.9942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.008826   0.023610  42.73  <2e-16 ***
## FirstAuthorFemale1 -0.017580   0.009339  -1.88   0.0598 .
## LastAuthorFemale1  0.011438   0.011054   1.03   0.3008
## UniqueAuthors2     0.188289   0.014728  12.78  <2e-16 ***
## UniqueAuthors3     0.222417   0.014955  14.87  <2e-16 ***
## UniqueAuthors4     0.251310   0.016893  14.88  <2e-16 ***
## UniqueAuthors5     0.325033   0.016472  19.73  <2e-16 ***
## Year1997           0.033522   0.029217   1.15   0.2513
## Year1998          -0.023580   0.030046  -0.78   0.4326
## Year1999          -0.005891   0.027823  -0.21   0.8323
```

```

## Year2000      -0.029978    0.028326    -1.06    0.2899
## Year2001      -0.029425    0.030685    -0.96    0.3376
## Year2002      -0.042896    0.028099    -1.53    0.1269
## Year2003      -0.075493    0.028348    -2.66    0.0078 **
## Year2004      -0.081079    0.029087    -2.79    0.0053 **
## Year2005      -0.055343    0.028640    -1.93    0.0533 .
## Year2006      -0.000173    0.028561    -0.01    0.9952
## Year2007      -0.028897    0.026537    -1.09    0.2762
## Year2008      -0.036486    0.027450    -1.33    0.1838
## Year2009      -0.019440    0.027546    -0.71    0.4804
## Year2010      -0.012981    0.027918    -0.46    0.6420
## Year2011      -0.010007    0.027208    -0.37    0.7130
## Year2012      -0.016707    0.027454    -0.61    0.5428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0657, Adjusted R-squared:  0.0635
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(6361,8357) are outliers with |weight| = 0 ( < 1.1e-05);
## 809 weights are ~ = 1. The remaining 8460 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0067 0.8580 0.9490 0.8930 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.08e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1 1.014
## LastAuthorFemale 1.016 1 1.008
## Year 1.031 16 1.001

```

## Residuals from first and last author



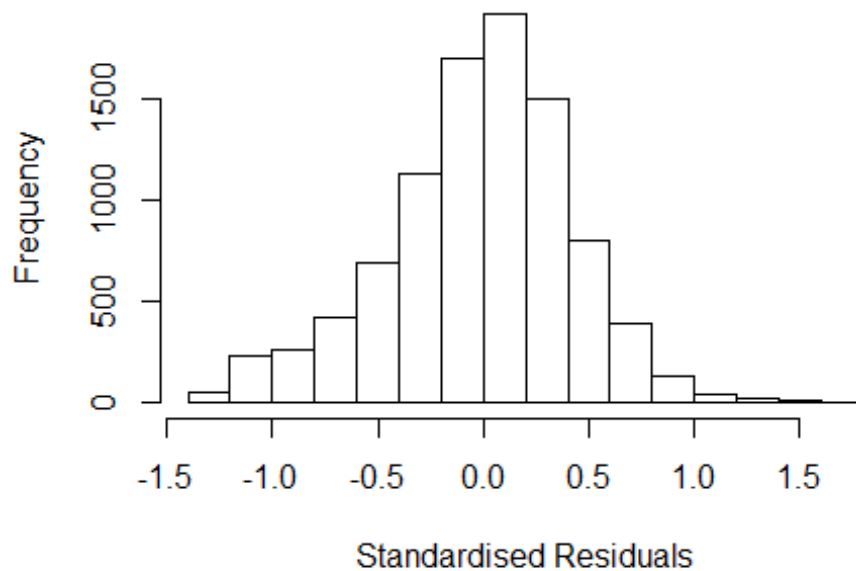
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2127 -0.2728 0.0156 0.2681 1.7915
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.138435 0.021025 54.15 <2e-16 ***
## FirstAuthorFemale1 0.008623 0.009409 0.92 0.359
## LastAuthorFemale1 0.000584 0.011131 0.05 0.958
## Year1997 0.043155 0.028598 1.51 0.131
## Year1998 -0.012316 0.029302 -0.42 0.674
## Year1999 0.020424 0.027368 0.75 0.456
## Year2000 0.000531 0.027952 0.02 0.985
## Year2001 0.016260 0.029916 0.54 0.587
## Year2002 0.006848 0.027451 0.25 0.803
## Year2003 -0.038758 0.028677 -1.35 0.177
## Year2004 -0.032053 0.028919 -1.11 0.268
## Year2005 -0.009740 0.028813 -0.34 0.735
```

```

## Year2006          0.061061    0.028225    2.16    0.031 *
## Year2007          0.042585    0.026059    1.63    0.102
## Year2008          0.025265    0.026907    0.94    0.348
## Year2009          0.049297    0.027231    1.81    0.070 .
## Year2010          0.060251    0.027310    2.21    0.027 *
## Year2011          0.063079    0.026776    2.36    0.019 *
## Year2012          0.065063    0.026826    2.43    0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.00635,    Adjusted R-squared:  0.00441
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 837 weights are ~= 1. The remaining 8434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0056 0.8590 0.9490 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year              1.024 16      1.001

```

## Residuals from first author



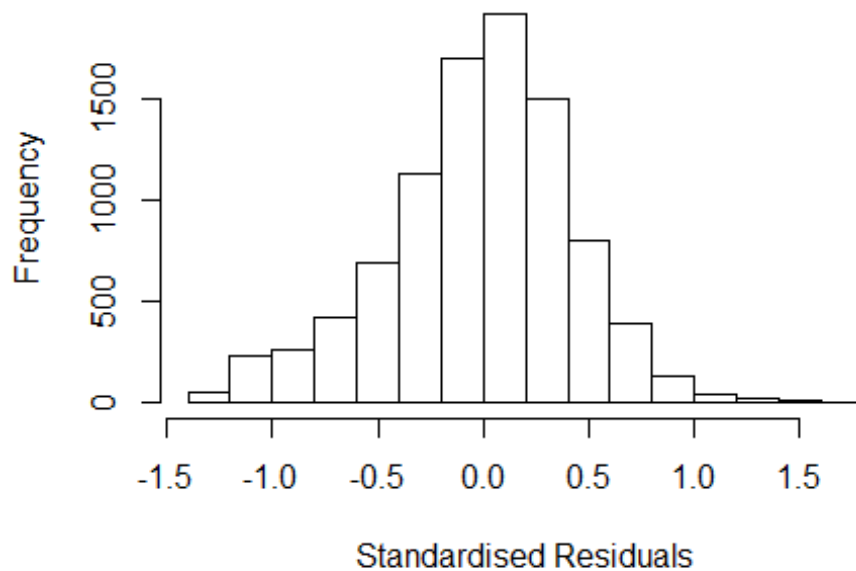
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2123 -0.2730 0.0155 0.2681 1.7914
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13854 0.02090 54.47 <2e-16 ***
## FirstAuthorFemale1 0.00872 0.00944 0.92 0.356
## Year1997 0.04313 0.02859 1.51 0.131
## Year1998 -0.01233 0.02930 -0.42 0.674
## Year1999 0.02040 0.02735 0.75 0.456
## Year2000 0.00052 0.02795 0.02 0.985
## Year2001 0.01625 0.02992 0.54 0.587
## Year2002 0.00685 0.02745 0.25 0.803
## Year2003 -0.03875 0.02868 -1.35 0.177
## Year2004 -0.03207 0.02891 -1.11 0.267
## Year2005 -0.00975 0.02881 -0.34 0.735
## Year2006 0.06105 0.02822 2.16 0.031 *
```

```

## Year2007          0.04258    0.02606    1.63    0.102
## Year2008          0.02526    0.02691    0.94    0.348
## Year2009          0.04929    0.02723    1.81    0.070 .
## Year2010          0.06025    0.02731    2.21    0.027 *
## Year2011          0.06307    0.02678    2.36    0.019 *
## Year2012          0.06505    0.02682    2.43    0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.00635,    Adjusted R-squared:  0.00452
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 837 weights are ~= 1. The remaining 8434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0055 0.8590 0.9490 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.011 1          1.005
## Year            1.011 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2089 -0.2724  0.0174  0.2681  1.7888
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14011    0.02087   54.63  <2e-16 ***
## LastAuthorFemale1 0.00247    0.01115    0.22   0.825
## Year1997        0.04345    0.02858    1.52   0.128
## Year1998       -0.01187    0.02929   -0.41   0.685
## Year1999        0.02041    0.02735    0.75   0.456
## Year2000        0.00066    0.02794    0.02   0.981
## Year2001        0.01651    0.02988    0.55   0.581
## Year2002        0.00752    0.02740    0.27   0.784
## Year2003       -0.03832    0.02865   -1.34   0.181
## Year2004       -0.03159    0.02890   -1.09   0.274
## Year2005       -0.00890    0.02878   -0.31   0.757
## Year2006        0.06174    0.02819    2.19   0.029 *
```

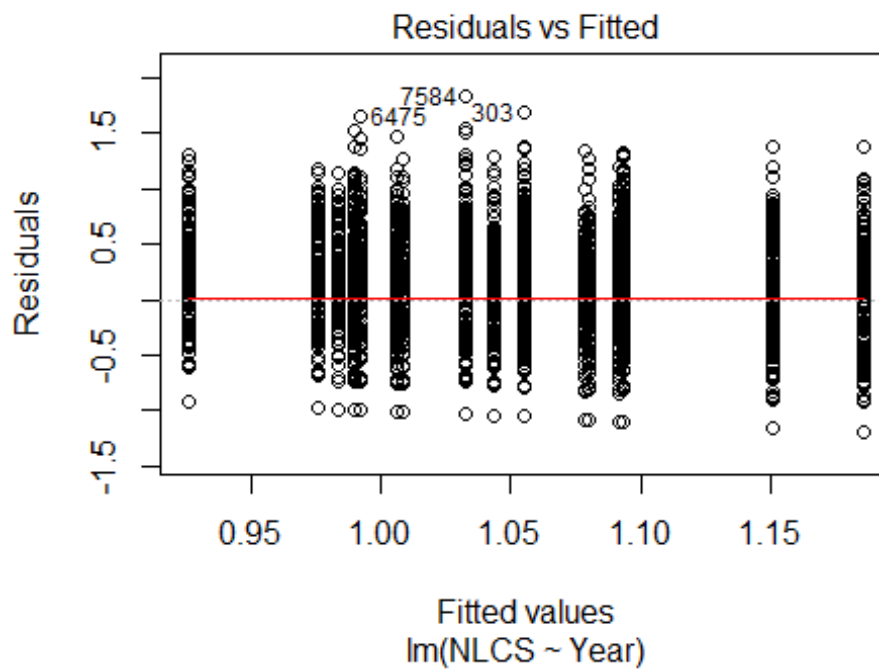


```

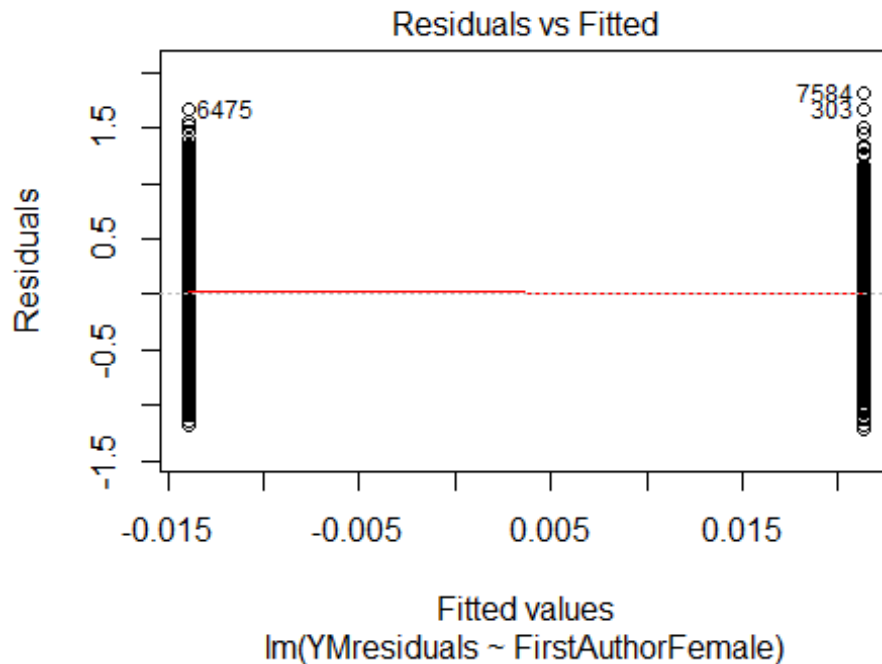
## Year2007      0.04333      0.02603      1.66      0.096 .
## Year2008      0.02618      0.02686      0.97      0.330
## Year2009      0.05047      0.02716      1.86      0.063 .
## Year2010      0.06127      0.02727      2.25      0.025 *
## Year2011      0.06413      0.02672      2.40      0.016 *
## Year2012      0.06631      0.02678      2.48      0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.00626,    Adjusted R-squared:  0.00444
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 817 weights are ~= 1. The remaining 8454 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.006  0.859  0.949  0.889  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 9271"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1106"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 464 447 472 462 504 460 433 387 373 279 305 344 326 301 299
## 2011 2012
## 336 302
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 216 197 193 202 212 164 179 190 150 140 160 186 192 179 177
## 2011 2012

```

```
## 210 196
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 199 177 174 184 185 149 164 171 132 118 139 162 175 162 168
## 2011 2012
## 193 183
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 55, df = 16, p-value = 4e-06
```

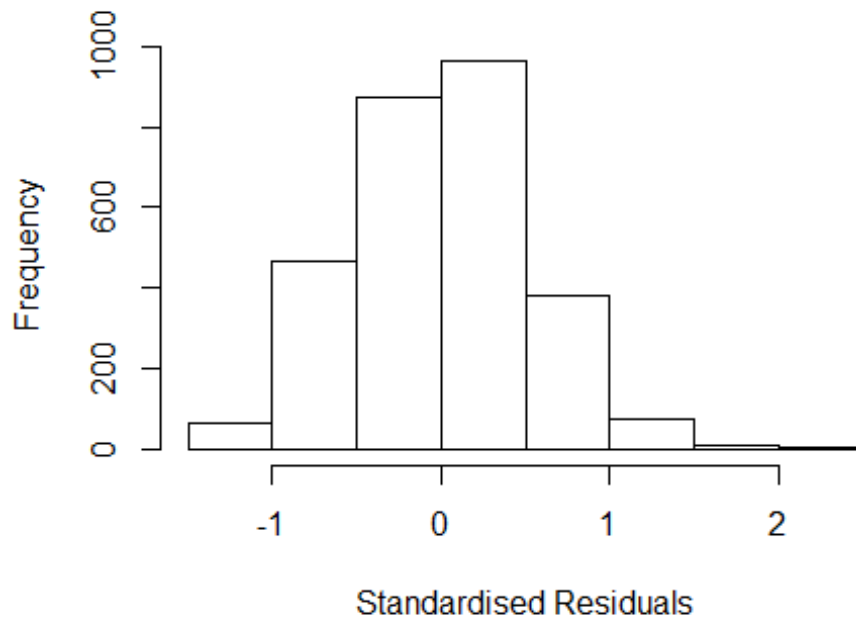


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.84343983070615"
## [1] "Male first author team size 2018 geometric mean: 2.91836108338131"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4800, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.71103400085464"
## [1] "Male last author team size 2018 geometric mean: 2.99861944483842"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1          1.043
## LastAuthorFemale  1.118 1          1.057
## UniqueAuthors    1.258 4          1.029
## Year              1.283 16         1.008
```

## Residuals from first and last author and team size



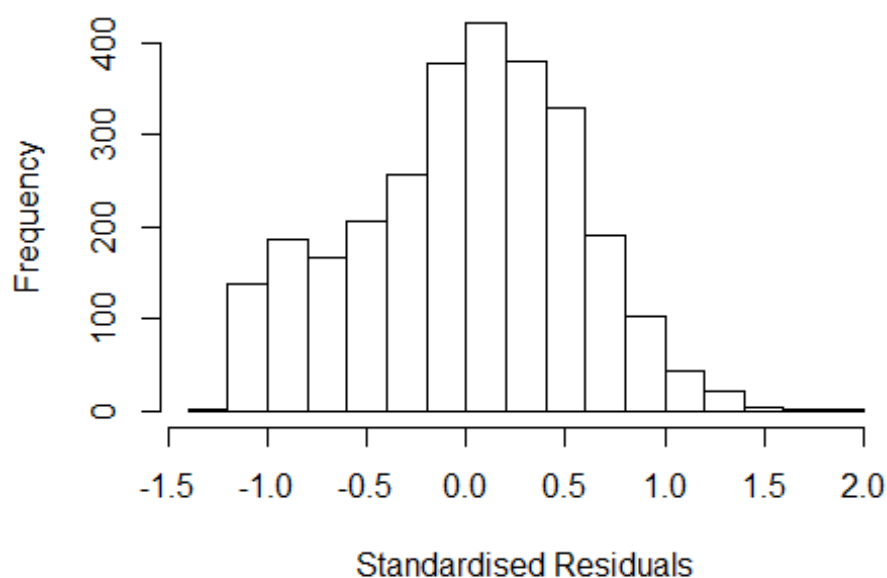
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.45394 -0.37076 0.00645 0.34078 2.36070
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.70845 0.04659 15.21 < 2e-16 ***
## FirstAuthorFemale1 0.04401 0.02066 2.13 0.03326 *
## LastAuthorFemale1 -0.09013 0.02413 -3.73 0.00019 ***
## UniqueAuthors2 0.43865 0.03623 12.11 < 2e-16 ***
## UniqueAuthors3 0.53931 0.03247 16.61 < 2e-16 ***
## UniqueAuthors4 0.59304 0.03448 17.20 < 2e-16 ***
## UniqueAuthors5 0.67584 0.03353 20.16 < 2e-16 ***
## Year1997 0.00462 0.06308 0.07 0.94165
## Year1998 0.06966 0.06076 1.15 0.25169
## Year1999 0.03397 0.05383 0.63 0.52796
```

```

## Year2000      -0.09458      0.05963      -1.59      0.11285
## Year2001      -0.05678      0.05808      -0.98      0.32838
## Year2002      -0.06656      0.05686      -1.17      0.24186
## Year2003      -0.03437      0.05360      -0.64      0.52138
## Year2004      -0.16524      0.05935      -2.78      0.00540 **
## Year2005      -0.07978      0.06208      -1.29      0.19889
## Year2006      -0.14068      0.05906      -2.38      0.01729 *
## Year2007      -0.07825      0.05718      -1.37      0.17125
## Year2008      -0.07790      0.05919      -1.32      0.18821
## Year2009      -0.14187      0.05997      -2.37      0.01807 *
## Year2010      -0.14357      0.06092      -2.36      0.01850 *
## Year2011      -0.15902      0.06681      -2.38      0.01738 *
## Year2012      -0.19741      0.05722      -3.45      0.00057 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.494
## Multiple R-squared:  0.218, Adjusted R-squared:  0.212
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2649 is an outlier with |weight| = 0 ( < 3.5e-05);
## 230 weights are ~= 1. The remaining 2604 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0356 0.8640 0.9450 0.9000 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## LastAuthorFemale 1.061 1      1.030
## Year      1.070 16      1.002

```

## Residuals from first and last author



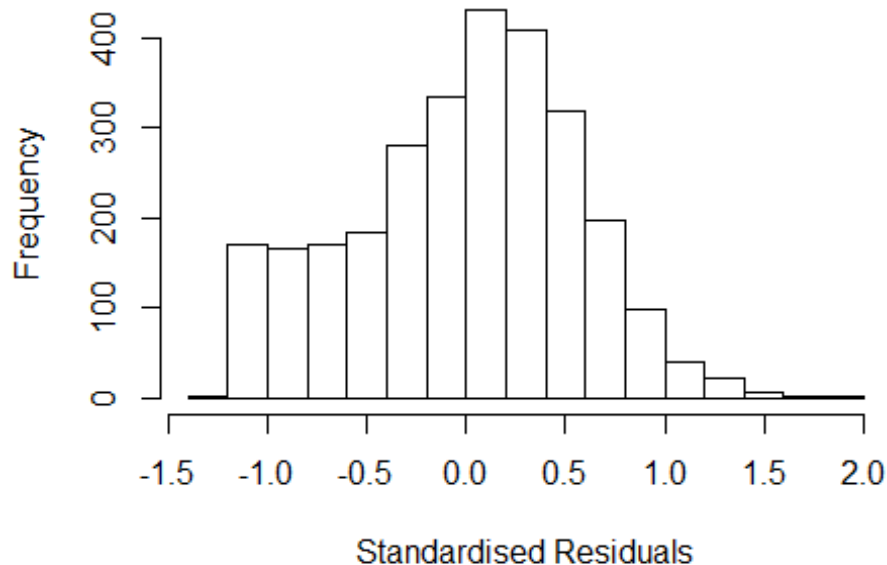
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2706 -0.3932 0.0399 0.3912 1.9219
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0628 0.0436 24.35 < 2e-16 ***
## FirstAuthorFemale1 0.0881 0.0225 3.92 9.1e-05 ***
## LastAuthorFemale1 -0.1704 0.0267 -6.38 2.0e-10 ***
## Year1997 0.0262 0.0676 0.39 0.698
## Year1998 0.1197 0.0631 1.90 0.058 .
## Year1999 0.1287 0.0567 2.27 0.023 *
## Year2000 -0.0633 0.0637 -0.99 0.320
## Year2001 -0.0122 0.0633 -0.19 0.847
## Year2002 -0.0420 0.0622 -0.68 0.499
## Year2003 0.0523 0.0587 0.89 0.373
## Year2004 -0.0791 0.0657 -1.20 0.229
## Year2005 0.0550 0.0617 0.89 0.373
```

```

## Year2006          -0.0233      0.0630   -0.37    0.711
## Year2007           0.0183      0.0589    0.31    0.756
## Year2008           0.0197      0.0637    0.31    0.757
## Year2009          -0.0677      0.0680   -1.00    0.319
## Year2010          -0.0736      0.0674   -1.09    0.274
## Year2011          -0.0384      0.0713   -0.54    0.590
## Year2012          -0.1134      0.0618   -1.83    0.067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.557
## Multiple R-squared:  0.0295, Adjusted R-squared:  0.0233
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 231 weights are ~= 1. The remaining 2604 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.209  0.858  0.947  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## Year              1.037 16          1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2104 -0.3863 0.0504 0.3910 1.8284
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03443 0.04481 23.09 <2e-16 ***
## FirstAuthorFemale1 0.04470 0.02306 1.94 0.053 .
## Year1997 0.03091 0.06917 0.45 0.655
## Year1998 0.11091 0.06450 1.72 0.086 .
## Year1999 0.13123 0.05846 2.24 0.025 *
## Year2000 -0.05885 0.06462 -0.91 0.362
## Year2001 -0.01508 0.06477 -0.23 0.816
## Year2002 -0.04072 0.06405 -0.64 0.525
## Year2003 0.05127 0.05996 0.85 0.393
## Year2004 -0.06830 0.06653 -1.03 0.305
## Year2005 0.06242 0.06254 1.00 0.318
## Year2006 -0.01796 0.06508 -0.28 0.783
```

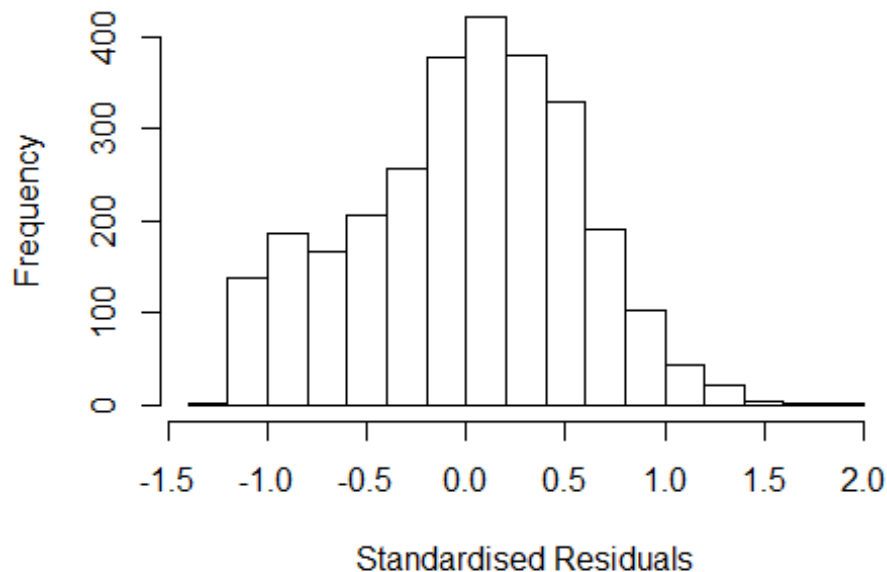


```

## Year2007          0.00203    0.06027    0.03    0.973
## Year2008          0.01927    0.06499    0.30    0.767
## Year2009         -0.06703    0.06960   -0.96    0.336
## Year2010         -0.07017    0.06779   -1.04    0.301
## Year2011         -0.04351    0.07200   -0.60    0.546
## Year2012         -0.11748    0.06310   -1.86    0.063 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.0145, Adjusted R-squared:  0.00856
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 224 weights are ~= 1. The remaining 2611 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.855  0.948  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.039 1          1.019
## Year            1.039 16          1.001

```

## Residuals from last author



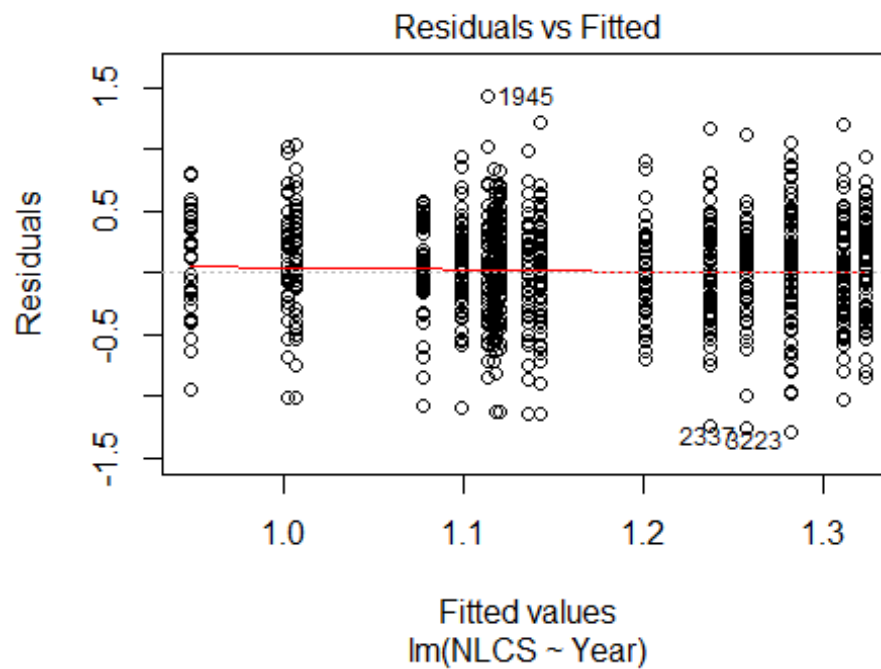
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2189 -0.3929 0.0475 0.3909 1.9527
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08736 0.04280 25.41 < 2e-16 ***
## LastAuthorFemale1 -0.14246 0.02645 -5.39 7.8e-08 ***
## Year1997 0.02464 0.06721 0.37 0.714
## Year1998 0.11949 0.06228 1.92 0.055 .
## Year1999 0.13152 0.05635 2.33 0.020 *
## Year2000 -0.06314 0.06338 -1.00 0.319
## Year2001 -0.00847 0.06302 -0.13 0.893
## Year2002 -0.03358 0.06175 -0.54 0.587
## Year2003 0.06120 0.05792 1.06 0.291
## Year2004 -0.07099 0.06459 -1.10 0.272
## Year2005 0.05926 0.06158 0.96 0.336
## Year2006 -0.01408 0.06221 -0.23 0.821
```

```

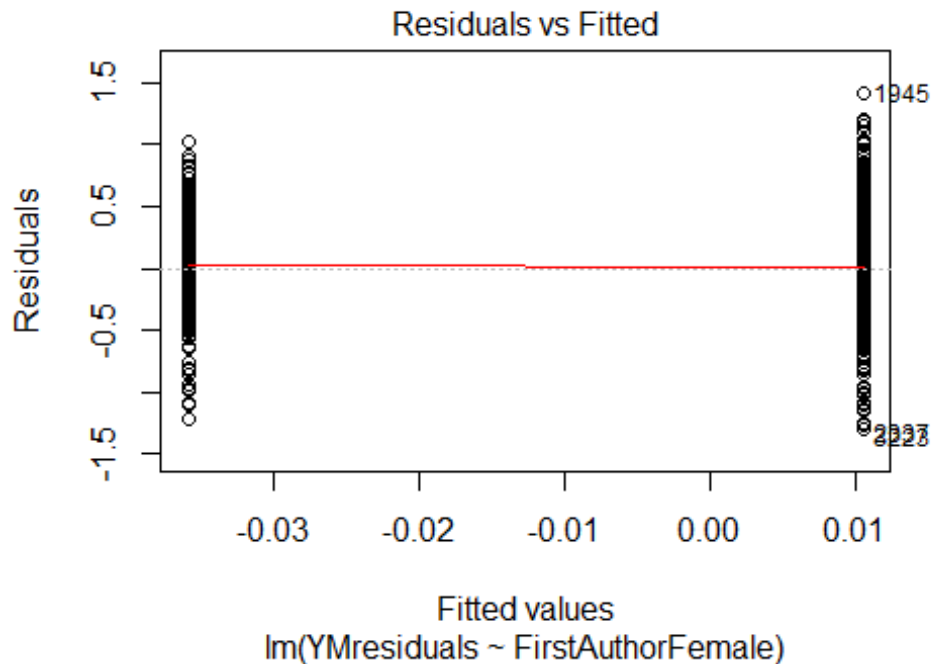
## Year2007      0.02879      0.05813      0.50      0.620
## Year2008      0.02314      0.06389      0.36      0.717
## Year2009     -0.06708      0.06822     -0.98      0.326
## Year2010     -0.07399      0.06708     -1.10      0.270
## Year2011     -0.03357      0.07109     -0.47      0.637
## Year2012     -0.11040      0.06210     -1.78      0.076 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0249, Adjusted R-squared:  0.019
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 235 weights are ~= 1. The remaining 2600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.193  0.857  0.947  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2835"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1107"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 162 169 125 149 154 148 128 121 113 91 91 126 107 135 127
## 2011 2012
## 142 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 61 36 54 72 47 47 46 44 43 46 65 60 70 71
## 2011 2012

```

```
## 75 75
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 55 34 53 64 41 44 42 42 41 41 57 54 66 68
## 2011 2012
## 67 69
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.09
```

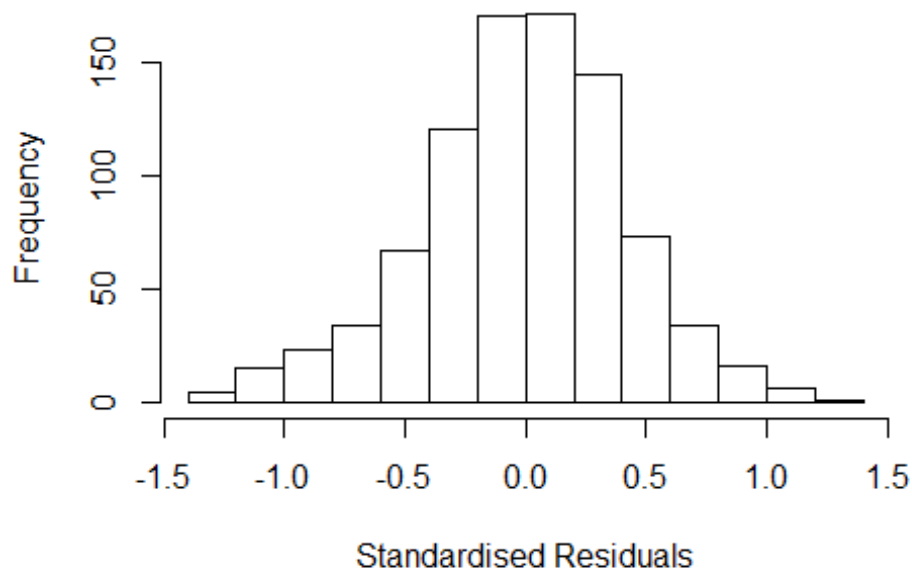


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.053, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 3.15245531536395"
## [1] "Male first author team size 2018 geometric mean: 3.21719026371197"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 840, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.04107934846993"
## [1] "Male last author team size 2018 geometric mean: 3.24898364883856"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 660, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.128 1          1.062
## LastAuthorFemale  1.148 1          1.071
## UniqueAuthors    1.620 4          1.062
## Year             1.962 16          1.021
```

## Residuals from first and last author and team size



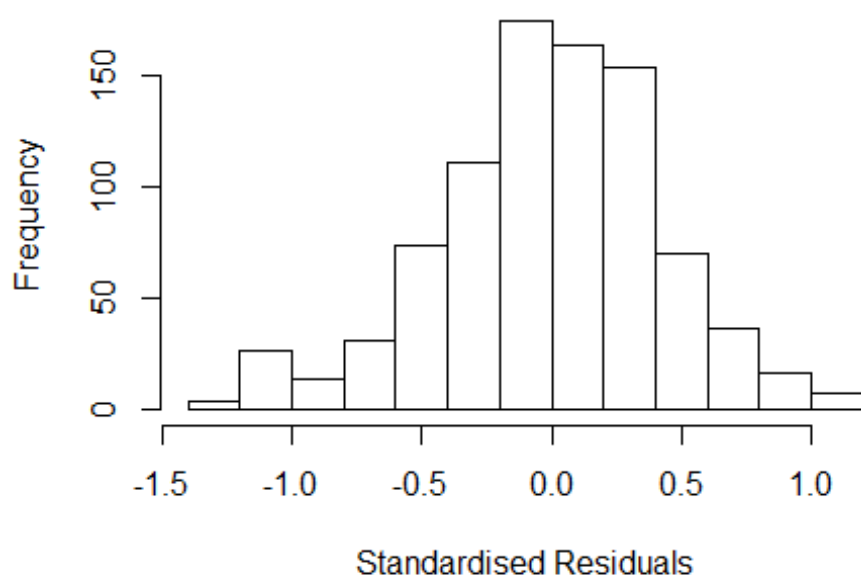
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29526 -0.26845  0.00616  0.25911  1.31663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93788    0.08457   11.09 < 2e-16 ***
## FirstAuthorFemale1 -0.02357    0.03477   -0.68  0.4981
## LastAuthorFemale1 -0.04233    0.04228   -1.00  0.3170
## UniqueAuthors2     0.19541    0.04588    4.26 2.3e-05 ***
## UniqueAuthors3     0.10993    0.04559    2.41  0.0161 *
## UniqueAuthors4     0.15192    0.04966    3.06  0.0023 **
## UniqueAuthors5     0.22988    0.05306    4.33 1.6e-05 ***
## Year1997          0.12117    0.09368    1.29  0.1962
## Year1998         -0.09112    0.11437   -0.80  0.4258
## Year1999          0.16197    0.10144    1.60  0.1107
```

```

## Year2000          0.09549    0.09285    1.03    0.3041
## Year2001          0.11608    0.11024    1.05    0.2927
## Year2002         -0.00837    0.11132   -0.08    0.9401
## Year2003          0.03784    0.09830    0.38    0.7004
## Year2004          0.06588    0.09485    0.69    0.4876
## Year2005          0.13869    0.09502    1.46    0.1448
## Year2006          0.02219    0.10073    0.22    0.8257
## Year2007          0.09364    0.10115    0.93    0.3548
## Year2008          0.17314    0.09264    1.87    0.0620 .
## Year2009          0.30101    0.09338    3.22    0.0013 **
## Year2010          0.15381    0.09212    1.67    0.0953 .
## Year2011          0.24867    0.09087    2.74    0.0063 **
## Year2012          0.21235    0.10260    2.07    0.0388 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0894, Adjusted R-squared:  0.0661
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 807 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.878  0.951  0.898  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.120 1      1.058
## LastAuthorFemale  1.120 1      1.058
## Year              1.226 16      1.006

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30027 -0.25693 0.00536 0.27240 1.18473
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0331 0.0822 12.56 <2e-16 ***
## FirstAuthorFemale1 -0.0164 0.0348 -0.47 0.6369
## LastAuthorFemale1 -0.0488 0.0422 -1.16 0.2478
## Year1997 0.0972 0.0965 1.01 0.3141
## Year1998 -0.0444 0.1191 -0.37 0.7092
## Year1999 0.1376 0.1046 1.32 0.1887
## Year2000 0.1322 0.0954 1.39 0.1662
## Year2001 0.1404 0.1116 1.26 0.2087
## Year2002 0.0203 0.1143 0.18 0.8593
## Year2003 0.0789 0.0995 0.79 0.4280
## Year2004 0.0943 0.0974 0.97 0.3332
## Year2005 0.1692 0.0984 1.72 0.0857 .
```

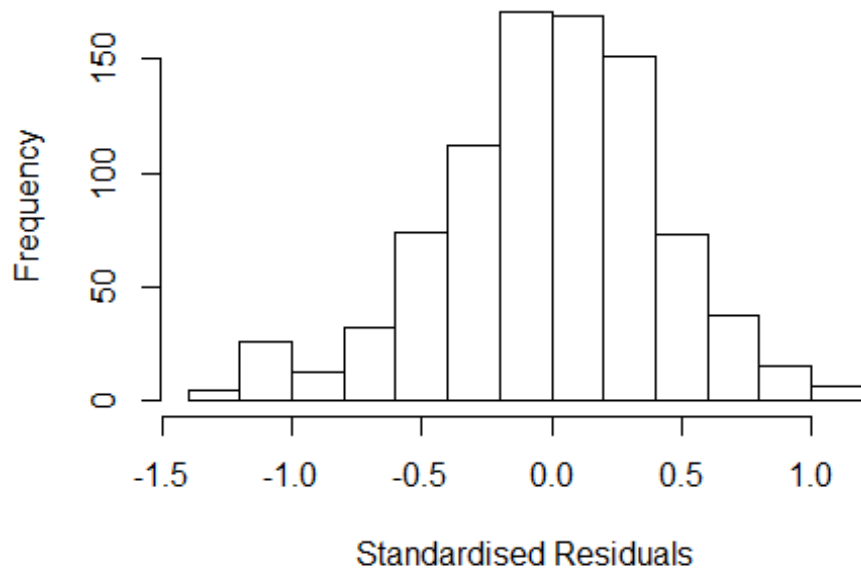


```

## Year2006          0.0573      0.1032      0.56      0.5786
## Year2007          0.1332      0.1016      1.31      0.1903
## Year2008          0.2451      0.0941      2.61      0.0093 **
## Year2009          0.3168      0.0951      3.33      0.0009 ***
## Year2010          0.2106      0.0938      2.25      0.0250 *
## Year2011          0.2922      0.0926      3.16      0.0017 **
## Year2012          0.2672      0.1047      2.55      0.0109 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.406
## Multiple R-squared:  0.0542, Adjusted R-squared:  0.0344
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 809 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.874   0.951   0.898   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.119 1      1.058
## Year              1.119 16      1.004

```

## Residuals from first author



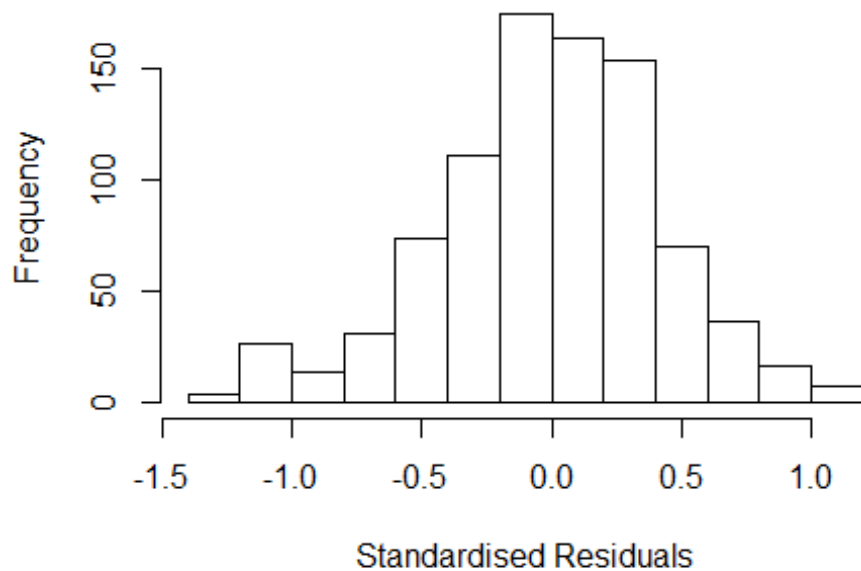
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29246 -0.26129 0.00993 0.27468 1.19227
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0272 0.0819 12.54 < 2e-16 ***
## FirstAuthorFemale1 -0.0242 0.0351 -0.69 0.48959
## Year1997 0.0953 0.0967 0.99 0.32452
## Year1998 -0.0390 0.1191 -0.33 0.74332
## Year1999 0.1406 0.1041 1.35 0.17725
## Year2000 0.1305 0.0956 1.36 0.17279
## Year2001 0.1454 0.1116 1.30 0.19305
## Year2002 0.0196 0.1140 0.17 0.86383
## Year2003 0.0758 0.0994 0.76 0.44565
## Year2004 0.0936 0.0974 0.96 0.33677
## Year2005 0.1745 0.0983 1.77 0.07638 .
## Year2006 0.0567 0.1037 0.55 0.58454
```

```

## Year2007          0.1359      0.1014      1.34  0.18058
## Year2008          0.2426      0.0943      2.57  0.01026 *
## Year2009          0.3175      0.0951      3.34  0.00088 ***
## Year2010          0.2112      0.0942      2.24  0.02520 *
## Year2011          0.2873      0.0928      3.10  0.00202 **
## Year2012          0.2652      0.1041      2.55  0.01104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0523, Adjusted R-squared:  0.0336
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 813 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.298  0.874  0.951  0.900  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.114 1      1.056
## Year              1.114 16      1.003

```

## Residuals from last author



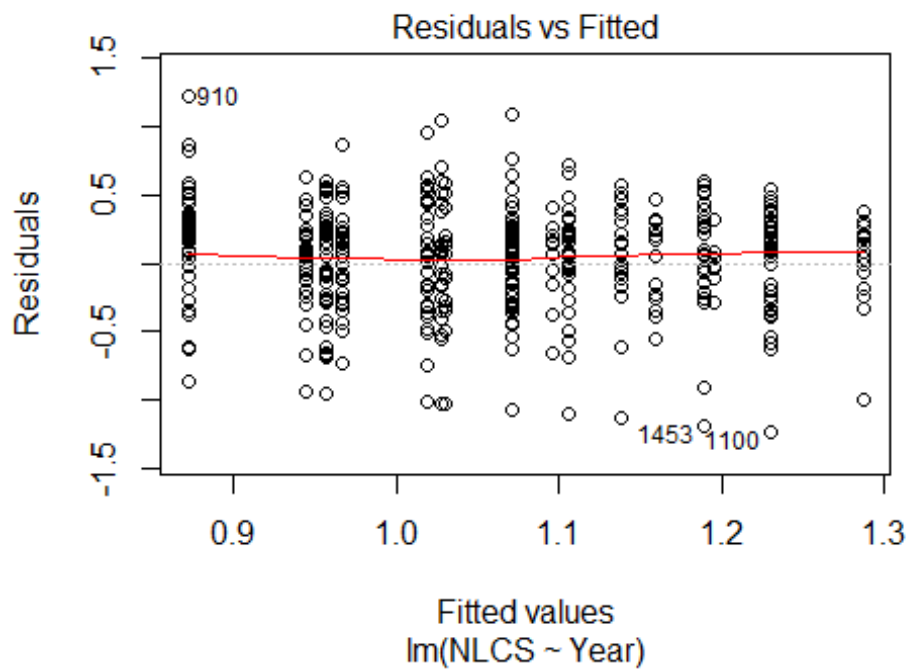
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29608 -0.25999  0.00558  0.27129  1.18791
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0310     0.0819   12.58 < 2e-16 ***
## LastAuthorFemale1 -0.0523     0.0422   -1.24  0.21612
## Year1997          0.0981     0.0966    1.02  0.30998
## Year1998         -0.0453     0.1194   -0.38  0.70462
## Year1999          0.1387     0.1044    1.33  0.18432
## Year2000          0.1311     0.0955    1.37  0.17003
## Year2001          0.1410     0.1116    1.26  0.20665
## Year2002          0.0215     0.1142    0.19  0.85085
## Year2003          0.0778     0.0998    0.78  0.43566
## Year2004          0.0938     0.0975    0.96  0.33631
## Year2005          0.1673     0.0988    1.69  0.09067 .
## Year2006          0.0568     0.1035    0.55  0.58357
```

```

## Year2007          0.1315      0.1017      1.29  0.19616
## Year2008          0.2428      0.0942      2.58  0.01011 *
## Year2009          0.3147      0.0952      3.31  0.00098 ***
## Year2010          0.2078      0.0937      2.22  0.02678 *
## Year2011          0.2906      0.0929      3.13  0.00181 **
## Year2012          0.2651      0.1047      2.53  0.01152 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0539, Adjusted R-squared:  0.0353
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 808 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.286  0.874  0.950  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 882"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1108"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 166 116 135 121 97 100 119 86 83 52 60 69 45 49 52
## 2011 2012
## 47 34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 34 37 29 7 31 55 35 16 10 21 27 15 16 21
## 2011 2012

```

```
## 23 20
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 30 30 23 5 22 39 34 14 10 18 25 15 14 21
## 2011 2012
## 23 20
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.004
```



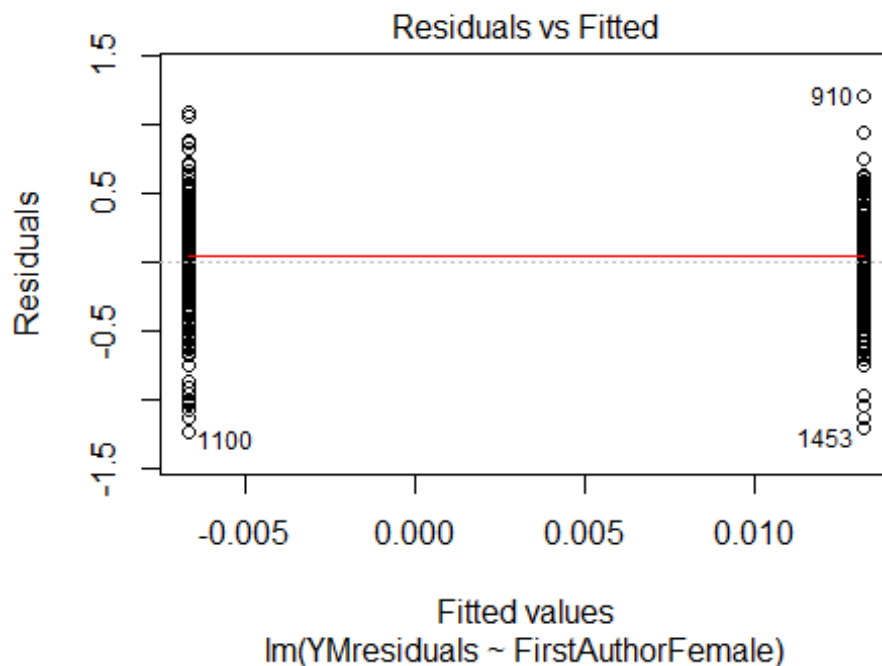
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.68, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 2.91369345857619"
## [1] "Male first author team size 2018 geometric mean: 1.8612097182042"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1"
## [1] "Male last author team size 2018 geometric mean: 2.36255439042876"

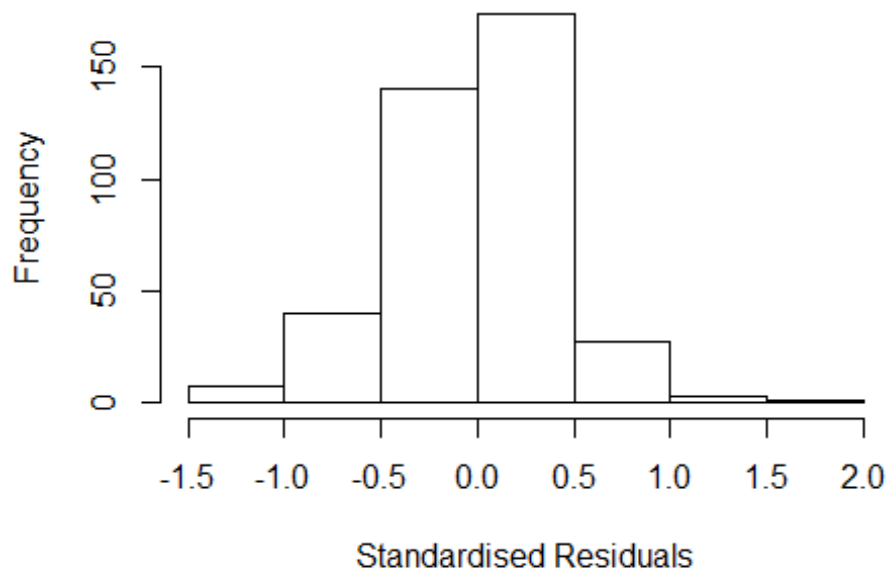
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.630	1	1.277
LastAuthorFemale	1.573	1	1.254
UniqueAuthors	3.082	4	1.151
Year	5.159	16	1.053

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3867 -0.2469 0.0194 0.2524 1.5151
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7629 0.1253 6.09 2.9e-09 ***
## FirstAuthorFemale1 0.0387 0.0490 0.79 0.43087
## LastAuthorFemale1 -0.0134 0.0669 -0.20 0.84115
## UniqueAuthors2 0.2827 0.1306 2.17 0.03101 *
## UniqueAuthors3 0.3391 0.1293 2.62 0.00910 **
## UniqueAuthors4 0.3995 0.1324 3.02 0.00272 **
## UniqueAuthors5 0.4979 0.1313 3.79 0.00017 ***
## Year1997 -0.0305 0.1255 -0.24 0.80791
## Year1998 -0.1662 0.0770 -2.16 0.03162 *
## Year1999 -0.1358 0.0909 -1.49 0.13631
```

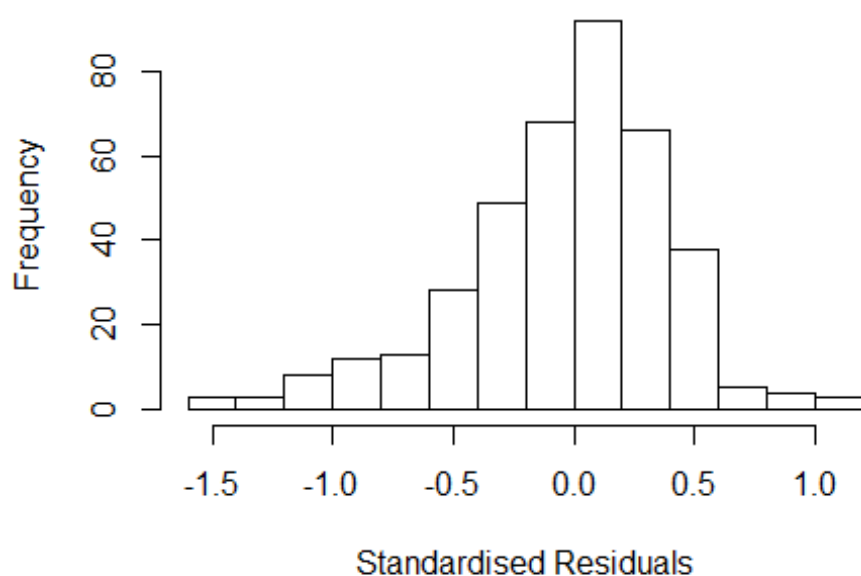


```

## Year2000          0.1516      0.0952      1.59  0.11223
## Year2001          0.0708      0.1017      0.70  0.48699
## Year2002         -0.2103      0.0977     -2.15  0.03205 *
## Year2003          0.0843      0.0916      0.92  0.35799
## Year2004         -0.0582      0.1333     -0.44  0.66237
## Year2005         -0.0395      0.1376     -0.29  0.77397
## Year2006         -0.1433      0.1124     -1.27  0.20328
## Year2007          0.2595      0.1127      2.30  0.02192 *
## Year2008          0.0488      0.0939      0.52  0.60377
## Year2009          0.1333      0.1012      1.32  0.18848
## Year2010          0.0176      0.1030      0.17  0.86420
## Year2011         -0.0807      0.1179     -0.68  0.49387
## Year2012         -0.1398      0.1044     -1.34  0.18149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.201, Adjusted R-squared:  0.154
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 363 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.061  0.857  0.955   0.894   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.701 1      1.304
## LastAuthorFemale  1.597 1      1.264
## Year              2.243 16      1.026

```

## Residuals from first and last author



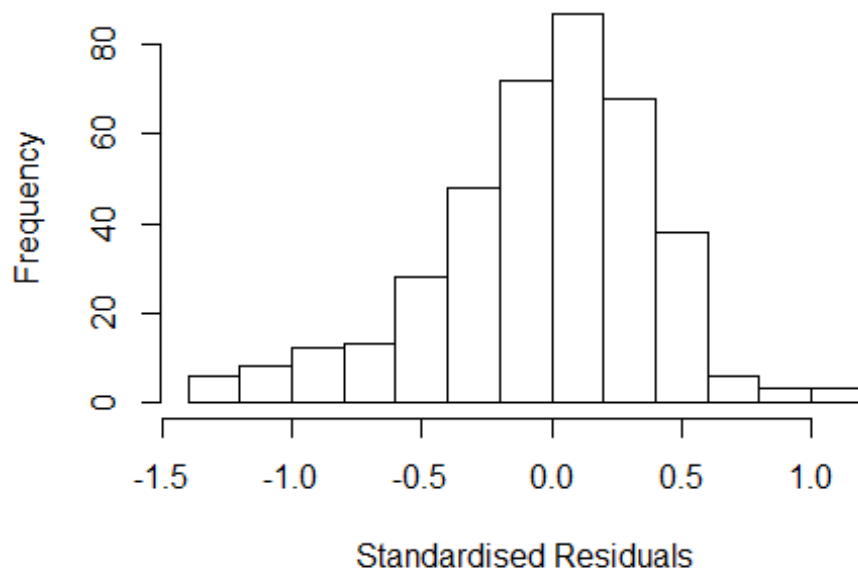
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.410 -0.252 0.029 0.252 1.125
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.053699 0.048725 21.63 <2e-16 ***
## FirstAuthorFemale1 0.063953 0.051133 1.25 0.2118
## LastAuthorFemale1 0.012600 0.068747 0.18 0.8547
## Year1997 -0.002116 0.110290 -0.02 0.9847
## Year1998 -0.101842 0.071794 -1.42 0.1569
## Year1999 -0.137503 0.093912 -1.46 0.1440
## Year2000 0.160835 0.093445 1.72 0.0860 .
## Year2001 0.055142 0.100936 0.55 0.5852
## Year2002 -0.162668 0.127330 -1.28 0.2022
## Year2003 0.165242 0.091893 1.80 0.0730 .
## Year2004 -0.000632 0.133090 0.00 0.9962
## Year2005 0.026687 0.115549 0.23 0.8175
```

```

## Year2006      -0.107876    0.117965   -0.91    0.3611
## Year2007      0.279651    0.101477    2.76    0.0061 **
## Year2008      0.094371    0.099610    0.95    0.3440
## Year2009      0.192494    0.101105    1.90    0.0577 .
## Year2010      0.150685    0.100067    1.51    0.1330
## Year2011     -0.034964    0.126399   -0.28    0.7822
## Year2012     -0.089602    0.110465   -0.81    0.4178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0996, Adjusted R-squared:  0.0561
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 365 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.135  0.869  0.952  0.888  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.503 1      1.226
## Year      1.503 16      1.013

```

## Residuals from first author



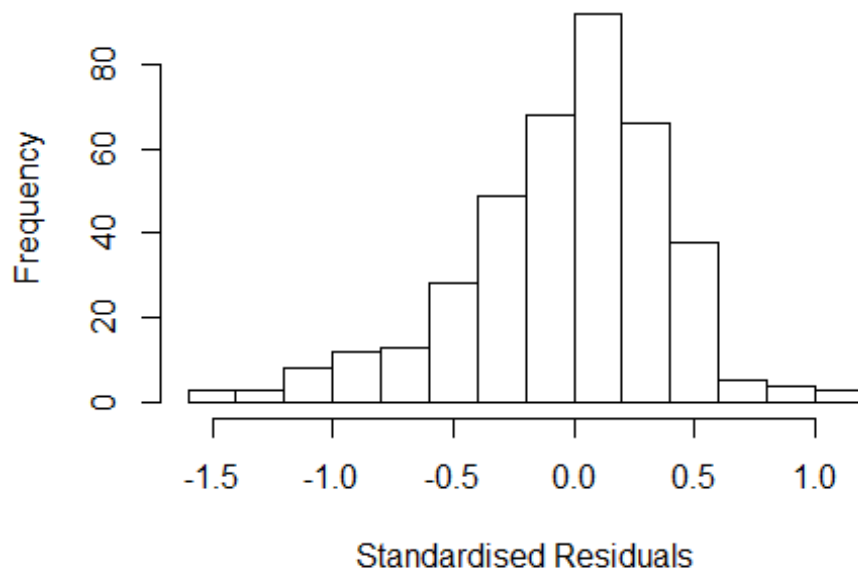
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3969 -0.2549 0.0273 0.2502 1.1367
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.054100 0.048953 21.53 <2e-16 ***
## FirstAuthorFemale1 0.063149 0.049861 1.27 0.2061
## Year1997 -0.000996 0.108574 -0.01 0.9927
## Year1998 -0.100547 0.070519 -1.43 0.1548
## Year1999 -0.134387 0.090761 -1.48 0.1395
## Year2000 0.165777 0.085284 1.94 0.0527 .
## Year2001 0.057176 0.098972 0.58 0.5638
## Year2002 -0.160937 0.126630 -1.27 0.2045
## Year2003 0.168473 0.087835 1.92 0.0559 .
## Year2004 0.000442 0.132398 0.00 0.9973
## Year2005 0.030042 0.112878 0.27 0.7903
## Year2006 -0.105953 0.116841 -0.91 0.3651
```

```

## Year2007          0.279638    0.101886    2.74    0.0064 **
## Year2008          0.096445    0.098952    0.97    0.3304
## Year2009          0.195723    0.097579    2.01    0.0456 *
## Year2010          0.154054    0.096450    1.60    0.1111
## Year2011         -0.033684    0.126369   -0.27    0.7900
## Year2012         -0.085811    0.106124   -0.81    0.4193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.0993, Adjusted R-squared:  0.0584
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 365 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.141  0.868  0.952  0.888  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.463 1          1.210
## Year            1.463 16          1.012

```

## Residuals from last author



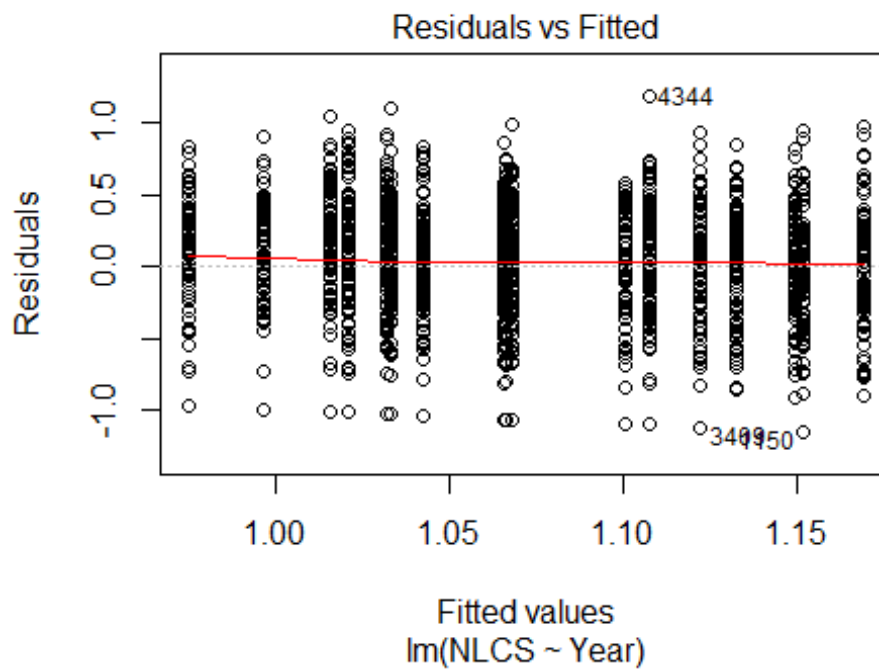
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3567 -0.2609 0.0355 0.2417 1.1687
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.066152 0.049063 21.73 <2e-16 ***
## LastAuthorFemale1 0.005639 0.067575 0.08 0.9335
## Year1997 -0.007685 0.111991 -0.07 0.9453
## Year1998 -0.105339 0.071762 -1.47 0.1430
## Year1999 -0.116656 0.087844 -1.33 0.1850
## Year2000 0.164451 0.090376 1.82 0.0696 .
## Year2001 0.074795 0.095216 0.79 0.4326
## Year2002 -0.147496 0.127983 -1.15 0.2499
## Year2003 0.188062 0.088250 2.13 0.0337 *
## Year2004 -0.000132 0.134997 0.00 0.9992
## Year2005 0.045367 0.110857 0.41 0.6826
## Year2006 -0.094666 0.121146 -0.78 0.4351
```

```

## Year2007          0.284898    0.104723    2.72    0.0068 **
## Year2008          0.096447    0.100112    0.96    0.3360
## Year2009          0.222312    0.091582    2.43    0.0157 *
## Year2010          0.159644    0.097654    1.63    0.1029
## Year2011         -0.018979    0.122419   -0.16    0.8769
## Year2012         -0.073729    0.110221   -0.67    0.5040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.377
## Multiple R-squared:  0.0938, Adjusted R-squared:  0.0526
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 359 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.169  0.869  0.950  0.885  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.55e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 392"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1109"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 238 236 212 209 247 248 259 179 200 191 145 183 179 158 164
## 2011 2012
## 173 179
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 110 91 89 87 94 64 146 100 112 92 90 110 114 113 111
## 2011 2012

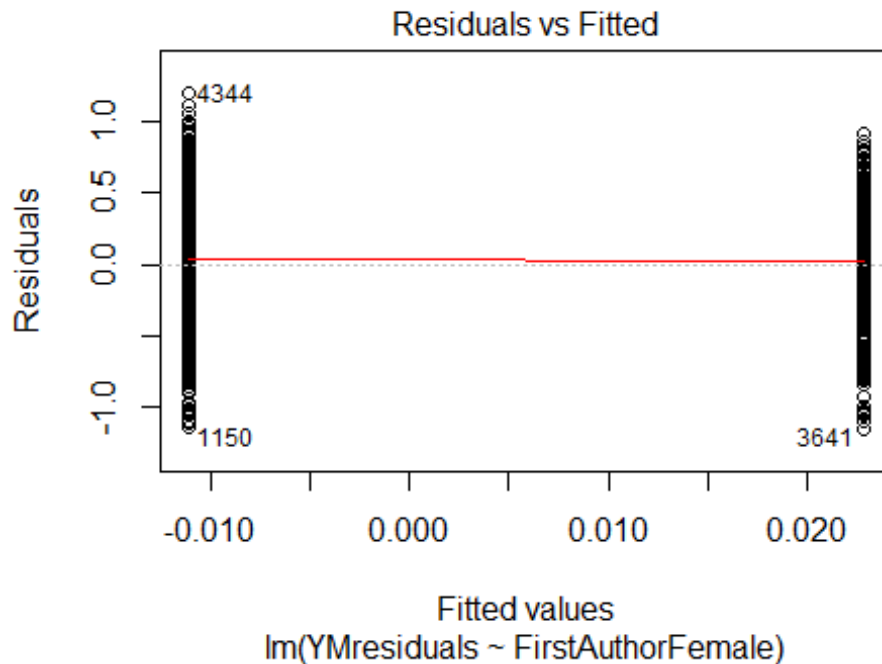
```

```
## 119 121
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 103 83 84 80 86 62 123 90 103 85 86 98 106 99 101
## 2011 2012
## 110 115
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 51, df = 16, p-value = 1e-05
```



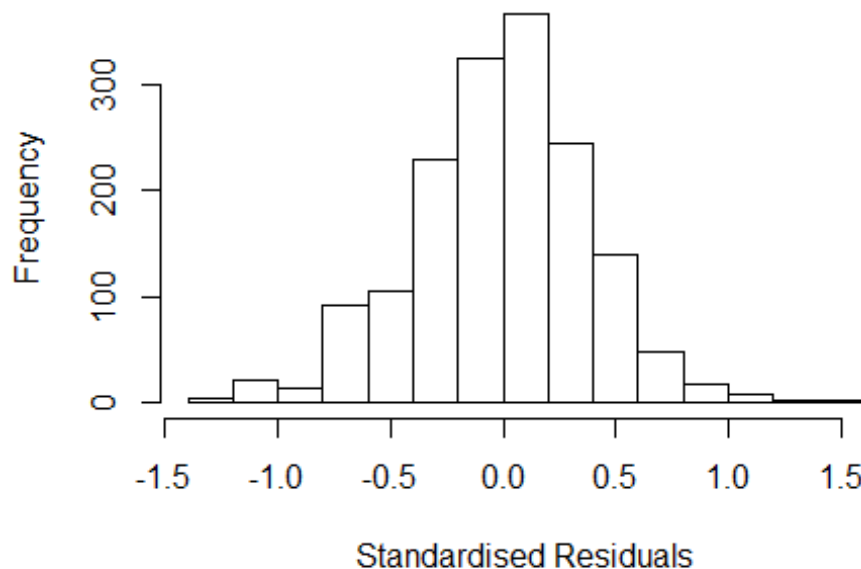
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 14, df = 1, p-value = 2e-04
```





```
## [1] "Female first author team size 2018 geometric mean: 3.9470684109254"
## [1] "Male first author team size 2018 geometric mean: 3.12592554242119"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 990, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.83400694621758"
## [1] "Male last author team size 2018 geometric mean: 3.27101825536601"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 770, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.073 1          1.036
## LastAuthorFemale  1.092 1          1.045
## UniqueAuthors    1.316 4          1.035
## Year             1.393 16          1.010
```

## Residuals from first and last author and team size



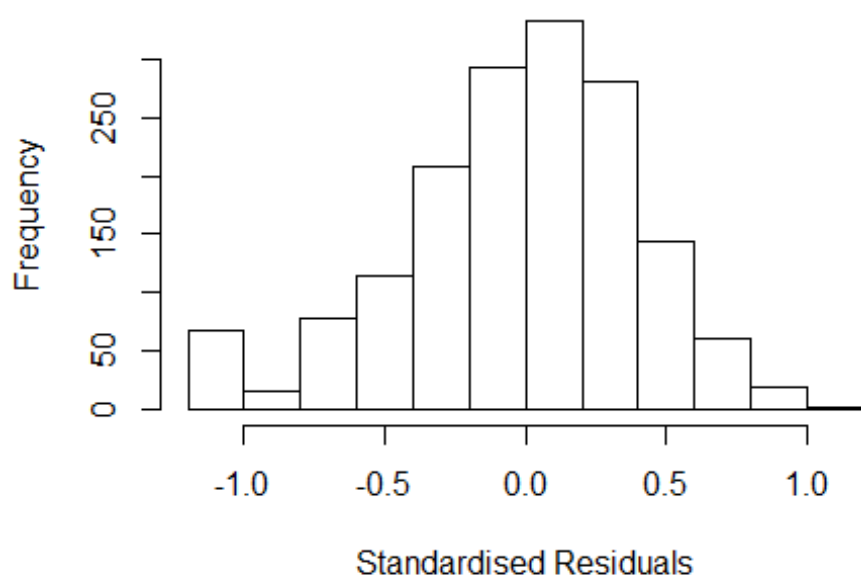
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3501 -0.2504 0.0166 0.2402 1.5621
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.73622 0.04775 15.42 <2e-16 ***
## FirstAuthorFemale1 -0.02331 0.01965 -1.19 0.236
## LastAuthorFemale1 -0.01513 0.02543 -0.60 0.552
## UniqueAuthors2 0.34334 0.03836 8.95 <2e-16 ***
## UniqueAuthors3 0.40530 0.03778 10.73 <2e-16 ***
## UniqueAuthors4 0.43412 0.03960 10.96 <2e-16 ***
## UniqueAuthors5 0.55766 0.03767 14.81 <2e-16 ***
## Year1997 0.13293 0.05857 2.27 0.023 *
## Year1998 0.07057 0.05345 1.32 0.187
## Year1999 0.11652 0.05643 2.06 0.039 *
```

```

## Year2000      0.03354      0.05419      0.62      0.536
## Year2001      0.05626      0.05951      0.95      0.345
## Year2002      0.01831      0.04793      0.38      0.702
## Year2003      0.00827      0.05220      0.16      0.874
## Year2004     -0.05301      0.05056     -1.05      0.295
## Year2005     -0.05492      0.05648     -0.97      0.331
## Year2006     -0.11148      0.05723     -1.95      0.052 .
## Year2007      0.00318      0.05168      0.06      0.951
## Year2008      0.03053      0.05101      0.60      0.550
## Year2009     -0.02550      0.05636     -0.45      0.651
## Year2010     -0.01757      0.06782     -0.26      0.796
## Year2011     -0.00432      0.05954     -0.07      0.942
## Year2012     -0.03489      0.05654     -0.62      0.537
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.192, Adjusted R-squared:  0.181
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1480 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0216 0.8660 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1      1.031
## LastAuthorFemale  1.057 1      1.028
## Year              1.117 16      1.003

```

## Residuals from first and last author



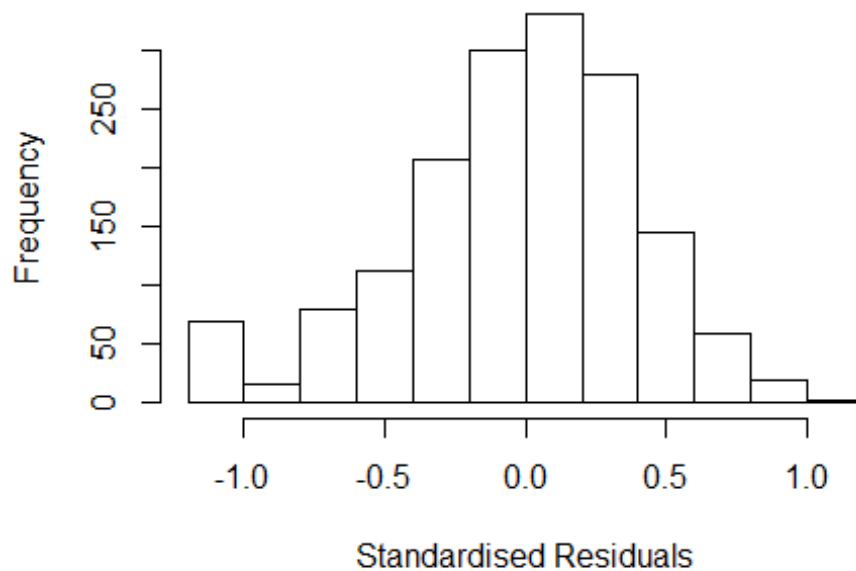
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1824 -0.2675  0.0192  0.2636  1.1462
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0358    0.0376   27.53  <2e-16 ***
## FirstAuthorFemale1  0.0212    0.0215    0.99   0.325
## LastAuthorFemale1 -0.0305    0.0277   -1.10   0.272
## Year1997          0.1457    0.0625    2.33   0.020 *
## Year1998          0.1269    0.0541    2.34   0.019 *
## Year1999          0.1145    0.0591    1.94   0.053 .
## Year2000          0.0282    0.0549    0.51   0.608
## Year2001          0.0912    0.0605    1.51   0.132
## Year2002          0.0714    0.0503    1.42   0.156
## Year2003          0.0610    0.0537    1.14   0.256
## Year2004          0.0131    0.0538    0.24   0.808
## Year2005          0.0123    0.0642    0.19   0.848
```

```

## Year2006          -0.0170      0.0586   -0.29    0.771
## Year2007           0.1001      0.0523    1.91    0.056 .
## Year2008           0.1255      0.0534    2.35    0.019 *
## Year2009           0.0190      0.0593    0.32    0.749
## Year2010           0.0309      0.0767    0.40    0.687
## Year2011           0.1121      0.0608    1.84    0.065 .
## Year2012           0.0334      0.0614    0.54    0.587
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0158, Adjusted R-squared:  0.00466
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1454 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.334  0.862  0.948  0.891  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.06 1      1.029
## Year              1.06 16      1.002

```

## Residuals from first author



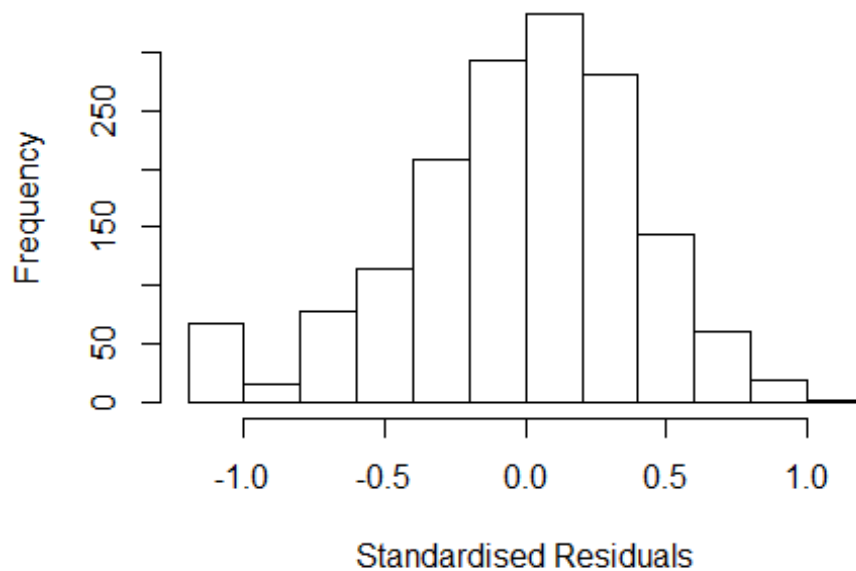
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1747 -0.2678 0.0185 0.2626 1.1517
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03402 0.03776 27.38 <2e-16 ***
## FirstAuthorFemale1 0.01754 0.02162 0.81 0.417
## Year1997 0.14456 0.06266 2.31 0.021 *
## Year1998 0.12335 0.05401 2.28 0.023 *
## Year1999 0.11400 0.05933 1.92 0.055 .
## Year2000 0.02409 0.05473 0.44 0.660
## Year2001 0.08920 0.06015 1.48 0.138
## Year2002 0.07012 0.05046 1.39 0.165
## Year2003 0.05861 0.05354 1.09 0.274
## Year2004 0.01146 0.05405 0.21 0.832
## Year2005 0.00511 0.06365 0.08 0.936
## Year2006 -0.01880 0.05854 -0.32 0.748
```

```

## Year2007          0.09939    0.05241    1.90    0.058 .
## Year2008          0.12312    0.05328    2.31    0.021 *
## Year2009          0.01584    0.05939    0.27    0.790
## Year2010          0.02796    0.07674    0.36    0.716
## Year2011          0.10826    0.06072    1.78    0.075 .
## Year2012          0.03130    0.06151    0.51    0.611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00455
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1463 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.340  0.861  0.948  0.891  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.057 1          1.028
## Year            1.057 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1696 -0.2660 0.0197 0.2633 1.1385
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0416 0.0372 28.01 <2e-16 ***
## LastAuthorFemale1 -0.0265 0.0277 -0.96 0.339
## Year1997 0.1465 0.0624 2.35 0.019 *
## Year1998 0.1283 0.0539 2.38 0.018 *
## Year1999 0.1158 0.0590 1.96 0.050 *
## Year2000 0.0280 0.0548 0.51 0.609
## Year2001 0.0930 0.0601 1.55 0.122
## Year2002 0.0719 0.0503 1.43 0.153
## Year2003 0.0600 0.0536 1.12 0.263
## Year2004 0.0155 0.0536 0.29 0.773
## Year2005 0.0117 0.0643 0.18 0.856
## Year2006 -0.0133 0.0584 -0.23 0.820
```

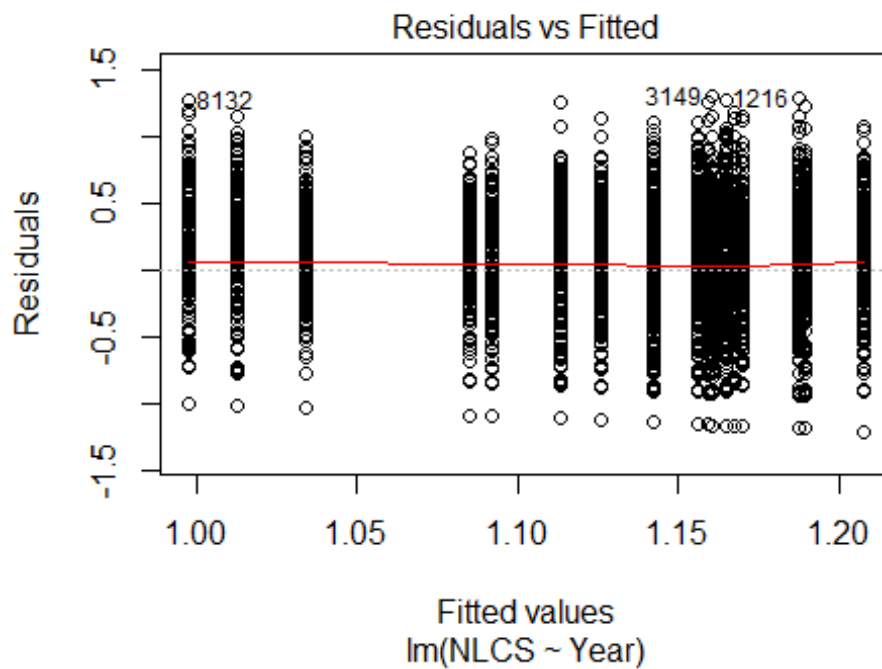


```

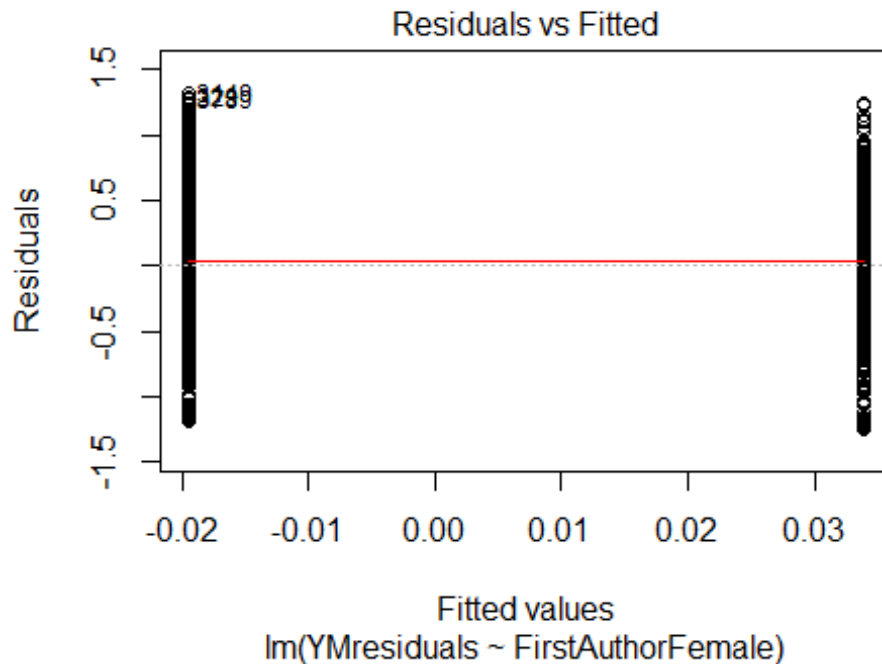
## Year2007          0.1007      0.0525      1.92      0.055 .
## Year2008          0.1280      0.0533      2.40      0.016 *
## Year2009          0.0181      0.0595      0.30      0.761
## Year2010          0.0329      0.0765      0.43      0.667
## Year2011          0.1138      0.0605      1.88      0.060 .
## Year2012          0.0361      0.0612      0.59      0.555
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.00471
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 152 weights are ~ = 1. The remaining 1462 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.344  0.863  0.948  0.891  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1614"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1110"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  833  699  741  625  687  592  529  443  424  384  377  373  401  357  336
## 2011 2012
##  332  326
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  335  278  338  305  222  177  316  255  209  226  226  236  260  239  231
## 2011 2012

```

```
## 221 241
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 303 255 301 278 204 157 274 239 194 208 210 222 240 224 219
## 2011 2012
## 200 221
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 48, df = 16, p-value = 4e-05
```

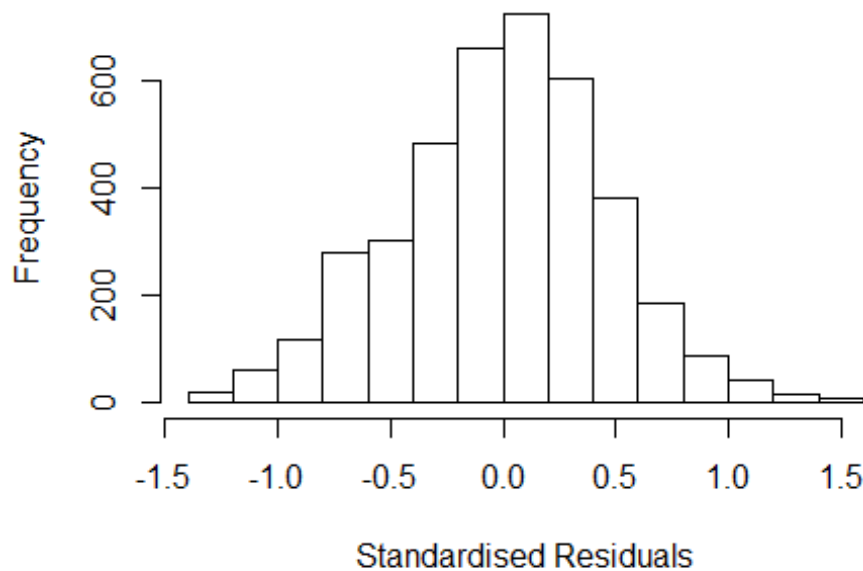


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 34, df = 1, p-value = 5e-09
```



```
## [1] "Female first author team size 2018 geometric mean: 3.49209555796617"
## [1] "Male first author team size 2018 geometric mean: 2.64620637387689"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5900, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.9565152171719"
## [1] "Male last author team size 2018 geometric mean: 2.97693439641392"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3900, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## LastAuthorFemale  1.043 1          1.021
## UniqueAuthors    1.180 4          1.021
## Year              1.215 16         1.006
```

## Residuals from first and last author and team size



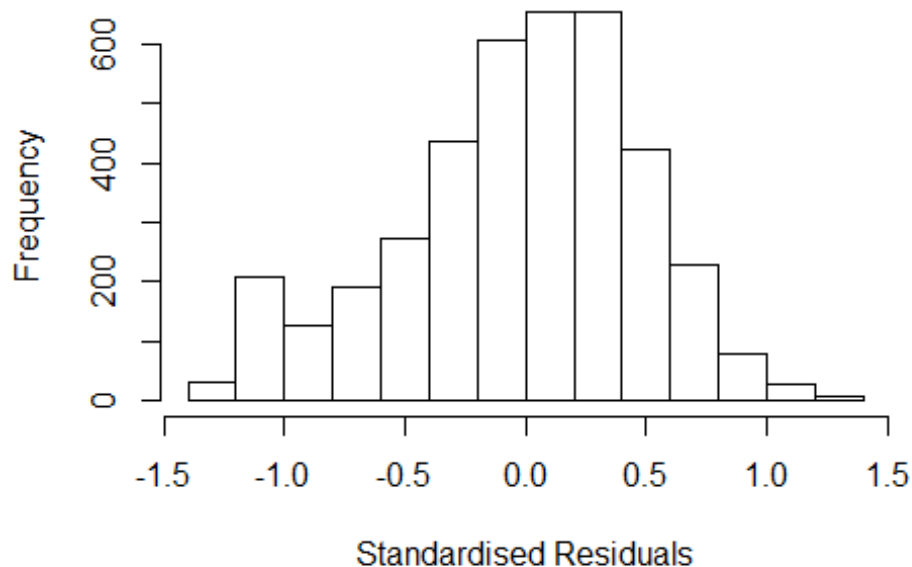
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3760 -0.2948 0.0198 0.2965 1.5786
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8494 0.0380 22.34 < 2e-16 ***
## FirstAuthorFemale1 0.0199 0.0149 1.33 0.18364
## LastAuthorFemale1 -0.0381 0.0183 -2.07 0.03812 *
## UniqueAuthors2 0.3512 0.0306 11.47 < 2e-16 ***
## UniqueAuthors3 0.4367 0.0294 14.88 < 2e-16 ***
## UniqueAuthors4 0.5248 0.0299 17.53 < 2e-16 ***
## UniqueAuthors5 0.6116 0.0294 20.77 < 2e-16 ***
## Year1997 -0.0224 0.0425 -0.53 0.59835
## Year1998 -0.0455 0.0407 -1.12 0.26460
## Year1999 -0.0638 0.0399 -1.60 0.10974
```

```

## Year2000          -0.0668      0.0458   -1.46   0.14464
## Year2001          -0.1691      0.0487   -3.48   0.00052 ***
## Year2002          -0.1047      0.0428   -2.45   0.01439 *
## Year2003          -0.1162      0.0413   -2.81   0.00491 **
## Year2004          -0.1663      0.0426   -3.90   9.6e-05 ***
## Year2005          -0.1910      0.0398   -4.80   1.6e-06 ***
## Year2006          -0.1871      0.0479   -3.90   9.7e-05 ***
## Year2007          -0.2295      0.0477   -4.81   1.6e-06 ***
## Year2008          -0.0949      0.0428   -2.22   0.02678 *
## Year2009          -0.0216      0.0450   -0.48   0.63119
## Year2010          -0.0587      0.0442   -1.33   0.18416
## Year2011          -0.1043      0.0449   -2.33   0.02009 *
## Year2012          -0.1306      0.0436   -3.00   0.00273 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.183, Adjusted R-squared:  0.179
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 327 weights are ~= 1. The remaining 3622 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.174  0.859  0.950  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.53e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.055 1          1.027
## LastAuthorFemale 1.029 1          1.014
## Year          1.083 16          1.003

```

## Residuals from first and last author



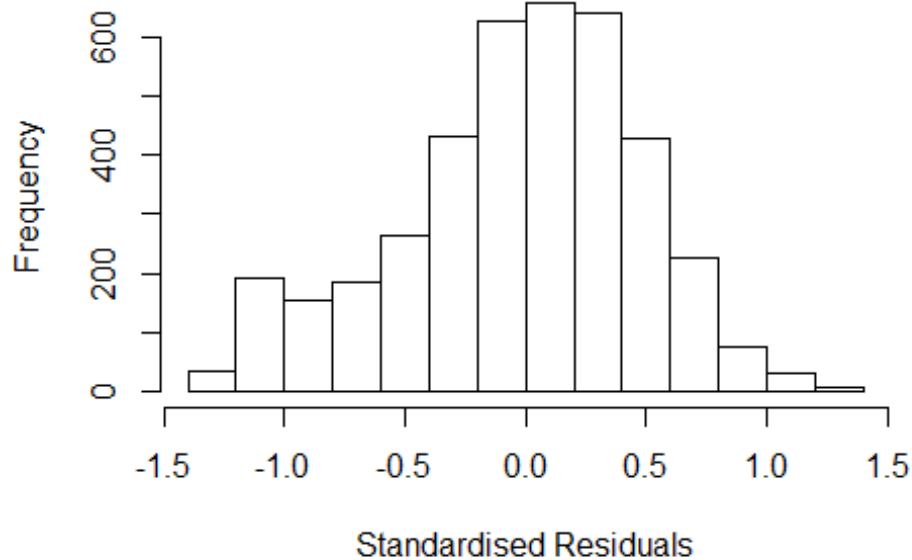
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2825 -0.3159 0.0286 0.3225 1.3783
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16815 0.03307 35.32 < 2e-16 ***
## FirstAuthorFemale1 0.06472 0.01627 3.98 7.0e-05 ***
## LastAuthorFemale1 -0.07946 0.02005 -3.96 7.5e-05 ***
## Year1997 0.03142 0.04469 0.70 0.4821
## Year1998 -0.00196 0.04320 -0.05 0.9638
## Year1999 -0.00995 0.04166 -0.24 0.8112
## Year2000 -0.01693 0.04607 -0.37 0.7133
## Year2001 -0.07462 0.04958 -1.51 0.1324
## Year2002 -0.01909 0.04450 -0.43 0.6679
## Year2003 -0.02853 0.04471 -0.64 0.5234
## Year2004 -0.05622 0.04590 -1.23 0.2206
## Year2005 -0.09994 0.04468 -2.24 0.0254 *
```

```

## Year2006          -0.12151      0.05654      -2.15      0.0317 *
## Year2007          -0.16715      0.05648      -2.96      0.0031 **
## Year2008          -0.03423      0.04752      -0.72      0.4713
## Year2009           0.04960      0.04913       1.01      0.3127
## Year2010           0.05570      0.04620       1.21      0.2281
## Year2011           0.00122      0.04976       0.02      0.9805
## Year2012           0.01035      0.04667       0.22      0.8245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.0142
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 333 weights are ~= 1. The remaining 3616 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.362  0.857   0.948   0.894   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.052 1      1.026
## Year              1.052 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2635 -0.3151 0.0263 0.3253 1.3120
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15635 0.03321 34.82 < 2e-16 ***
## FirstAuthorFemale1 0.05474 0.01660 3.30 0.00098 ***
## Year1997 0.03634 0.04498 0.81 0.41912
## Year1998 -0.00469 0.04339 -0.11 0.91386
## Year1999 -0.01140 0.04203 -0.27 0.78627
## Year2000 -0.01438 0.04653 -0.31 0.75721
## Year2001 -0.07381 0.05013 -1.47 0.14100
## Year2002 -0.02007 0.04481 -0.45 0.65430
## Year2003 -0.02945 0.04471 -0.66 0.51018
## Year2004 -0.05551 0.04637 -1.20 0.23130
## Year2005 -0.10066 0.04503 -2.24 0.02544 *
## Year2006 -0.12622 0.05674 -2.22 0.02617 *
```

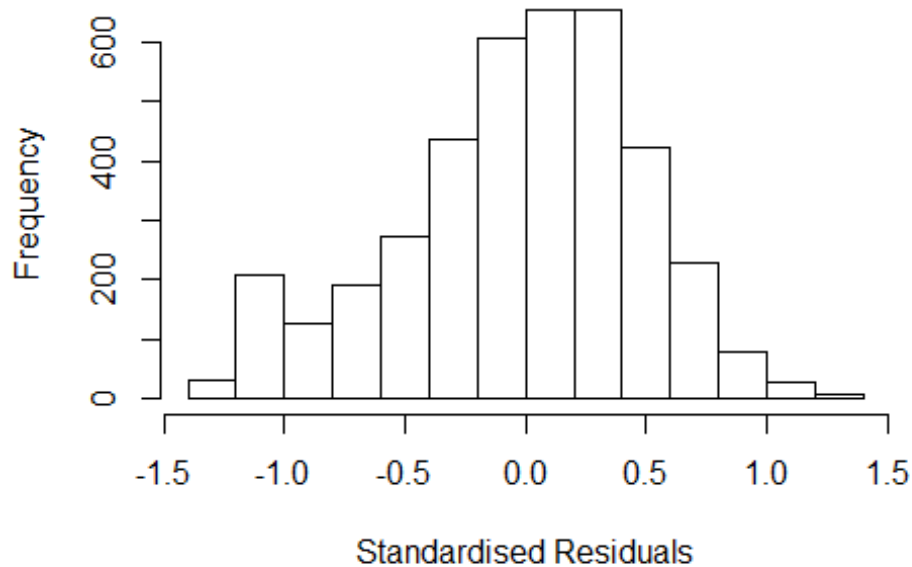


```

## Year2007          -0.16452    0.05709   -2.88  0.00398 **
## Year2008          -0.03733    0.04760   -0.78  0.43300
## Year2009           0.05073    0.04931    1.03  0.30362
## Year2010           0.05237    0.04650    1.13  0.26014
## Year2011          -0.00348    0.04982   -0.07  0.94424
## Year2012           0.00840    0.04696    0.18  0.85796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.0108
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 340 weights are ~= 1. The remaining 3609 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.410  0.855  0.947  0.893  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.015
## Year              1.029 16          1.001

```

## Residuals from last author



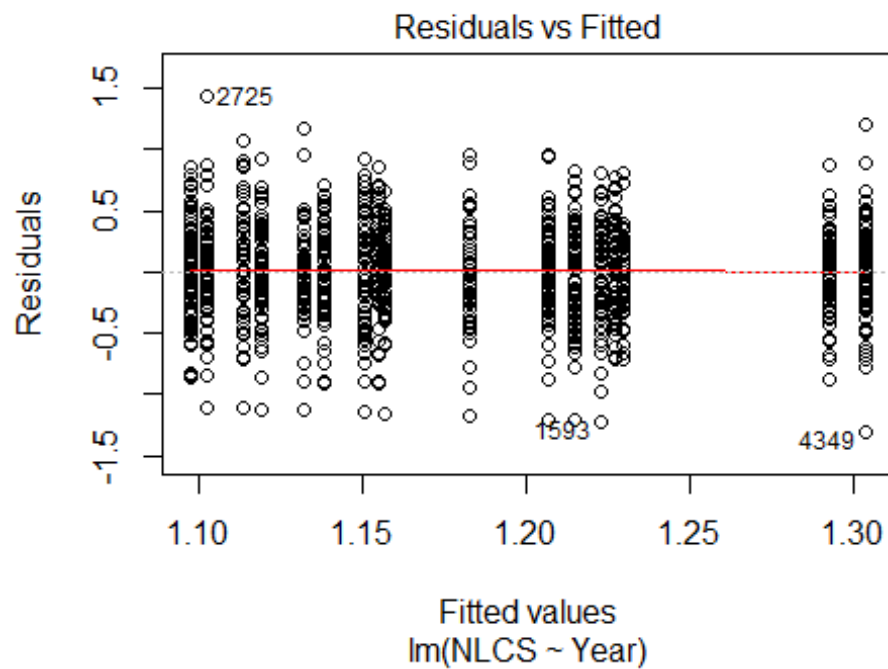
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2494 -0.3160 0.0321 0.3277 1.3423
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18735 0.03245 36.58 < 2e-16 ***
## LastAuthorFemale1 -0.06739 0.02018 -3.34 0.00085 ***
## Year1997 0.02948 0.04479 0.66 0.51050
## Year1998 -0.00369 0.04317 -0.09 0.93195
## Year1999 -0.00524 0.04159 -0.13 0.89977
## Year2000 -0.01559 0.04591 -0.34 0.73414
## Year2001 -0.06931 0.04953 -1.40 0.16185
## Year2002 -0.01615 0.04460 -0.36 0.71720
## Year2003 -0.01905 0.04469 -0.43 0.66986
## Year2004 -0.05665 0.04599 -1.23 0.21809
## Year2005 -0.09274 0.04466 -2.08 0.03793 *
## Year2006 -0.11406 0.05684 -2.01 0.04486 *
```

```

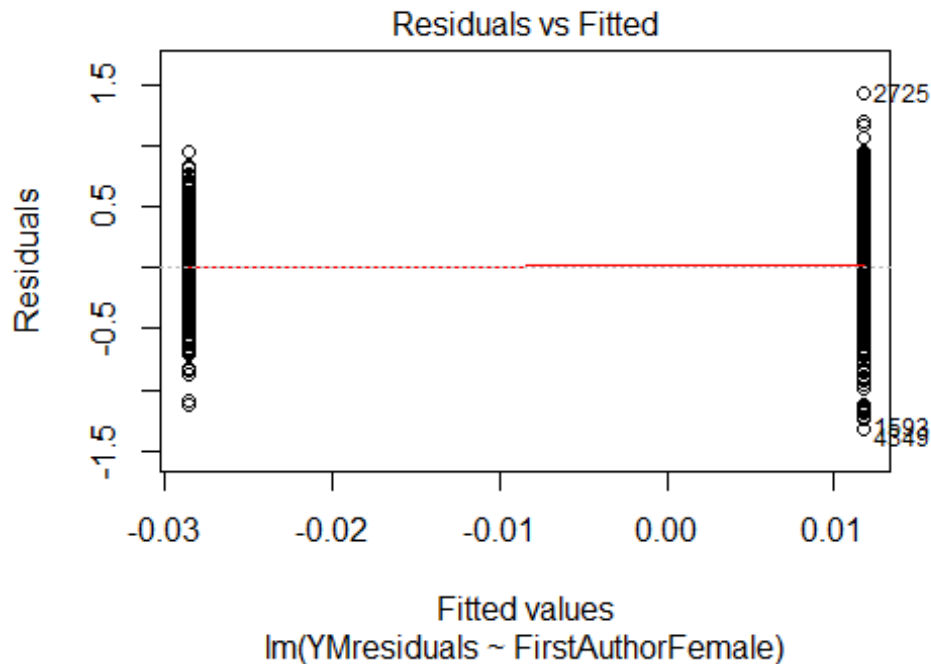
## Year2007          -0.15896      0.05659    -2.81  0.00500 **
## Year2008          -0.03084      0.04756    -0.65  0.51681
## Year2009           0.05419      0.04891     1.11  0.26799
## Year2010           0.06203      0.04619     1.34  0.17936
## Year2011           0.00499      0.04991     0.10  0.92040
## Year2012           0.01827      0.04651     0.39  0.69442
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.015, Adjusted R-squared:  0.0107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 325 weights are ~= 1. The remaining 3624 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.387  0.857  0.947  0.893  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3949"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1111"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 219 205 198 204 207 206 175 172 139 144 172 196 145 142 163
## 2011 2012
## 176 155
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 47 53 62 78 66 59 51 51 56 70 77 66 59 73
## 2011 2012

```

```
## 74 68
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 40 47 58 72 62 49 49 47 54 62 64 61 51 66
## 2011 2012
## 65 62
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.07
```

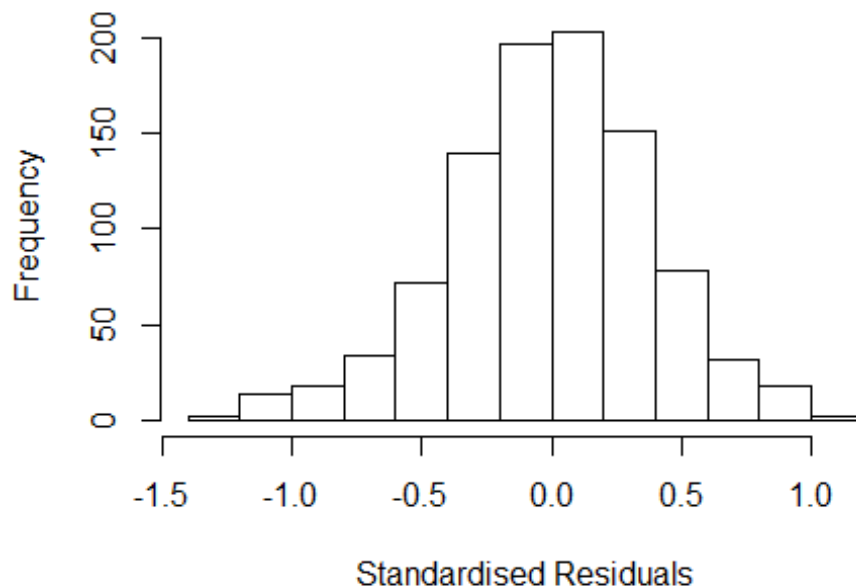


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.3, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 4.85452318295509"
## [1] "Male first author team size 2018 geometric mean: 3.22072194572276"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 700, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.98493617990518"
## [1] "Male last author team size 2018 geometric mean: 3.42785821141657"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 390, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.162 1      1.078
## LastAuthorFemale  1.126 1      1.061
## UniqueAuthors    1.549 4      1.056
## Year              1.704 16     1.017
```

## Residuals from first and last author and team size



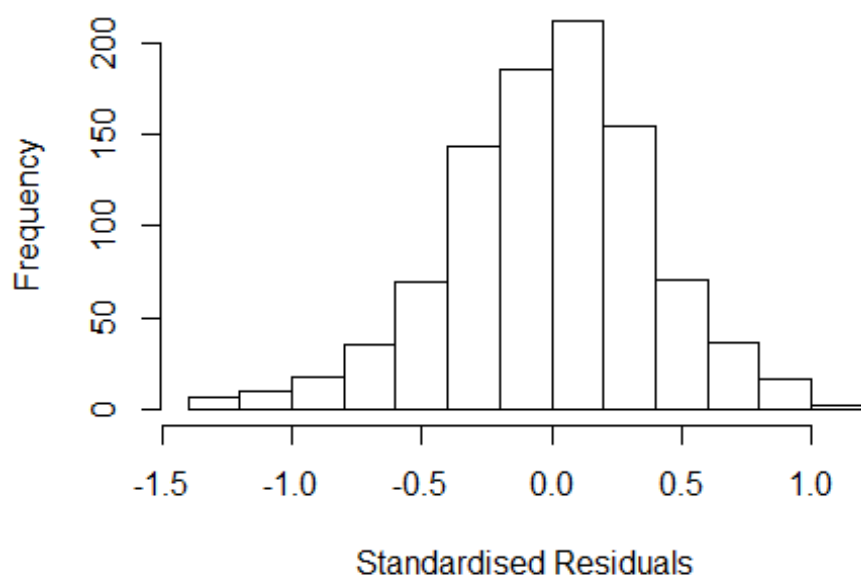
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30145 -0.24908 0.00841 0.25006 1.15682
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10560 0.07984 13.85 <2e-16 ***
## FirstAuthorFemale1 -0.03927 0.02848 -1.38 0.1682
## LastAuthorFemale1 -0.04403 0.03616 -1.22 0.2237
## UniqueAuthors2 0.06183 0.05124 1.21 0.2279
## UniqueAuthors3 0.03802 0.05173 0.73 0.4626
## UniqueAuthors4 0.08599 0.05241 1.64 0.1012
## UniqueAuthors5 0.13939 0.05318 2.62 0.0089 **
## Year1997 0.10453 0.09024 1.16 0.2470
## Year1998 0.06860 0.09405 0.73 0.4660
## Year1999 0.00554 0.09899 0.06 0.9554
```

```

## Year2000      -0.00635    0.08613   -0.07    0.9413
## Year2001      0.10005    0.09247    1.08    0.2795
## Year2002      0.02319    0.09371    0.25    0.8046
## Year2003     -0.02424    0.09124   -0.27    0.7905
## Year2004      0.02047    0.08619    0.24    0.8123
## Year2005      0.01350    0.09413    0.14    0.8860
## Year2006     -0.04346    0.09068   -0.48    0.6319
## Year2007     -0.03307    0.09017   -0.37    0.7139
## Year2008      0.12000    0.08674    1.38    0.1669
## Year2009      0.06836    0.09366    0.73    0.4656
## Year2010      0.06603    0.08873    0.74    0.4569
## Year2011      0.03438    0.09186    0.37    0.7083
## Year2012      0.13402    0.09167    1.46    0.1441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.0332, Adjusted R-squared:  0.0105
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 881 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.207  0.873   0.952   0.900   0.986   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           1.04e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.130 1           1.063
## LastAuthorFemale  1.110 1           1.053
## Year              1.167 16           1.005

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31474 -0.25701 0.00969 0.25257 1.15933
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14843 0.07441 15.43 <2e-16 ***
## FirstAuthorFemale1 -0.03004 0.02816 -1.07 0.286
## LastAuthorFemale1 -0.03802 0.03619 -1.05 0.294
## Year1997 0.10056 0.09107 1.10 0.270
## Year1998 0.08467 0.09309 0.91 0.363
## Year1999 0.00374 0.09991 0.04 0.970
## Year2000 0.01521 0.08609 0.18 0.860
## Year2001 0.11427 0.09233 1.24 0.216
## Year2002 0.03861 0.09292 0.42 0.678
## Year2003 -0.00775 0.09029 -0.09 0.932
## Year2004 0.05300 0.08602 0.62 0.538
## Year2005 0.02758 0.09508 0.29 0.772
```

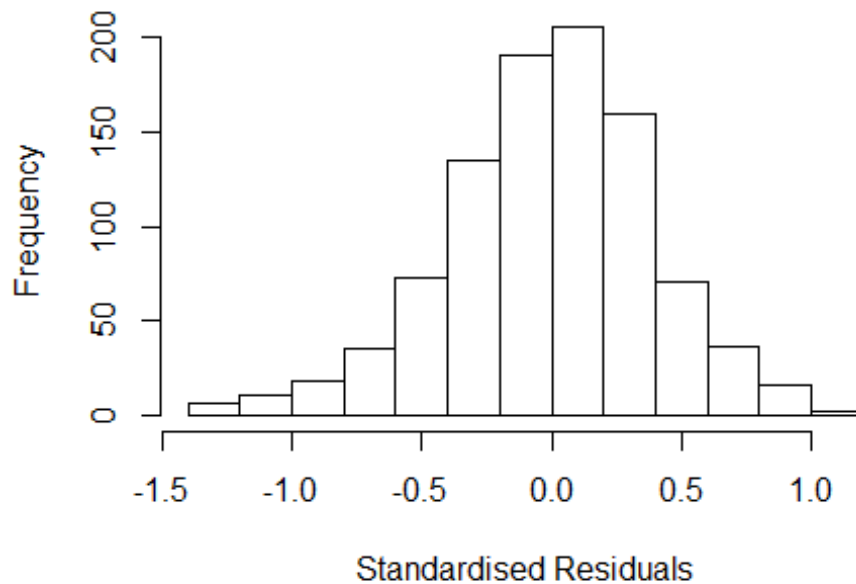


```

## Year2006      -0.01295    0.08863   -0.15    0.884
## Year2007      -0.01563    0.09013   -0.17    0.862
## Year2008       0.14908    0.08555    1.74    0.082 .
## Year2009       0.10143    0.09274    1.09    0.274
## Year2010       0.09093    0.08840    1.03    0.304
## Year2011       0.07403    0.08971    0.83    0.409
## Year2012       0.16632    0.09037    1.84    0.066 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.00405
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 883 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.872  0.951  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.097 1      1.047
## Year              1.097 16      1.003

```

## Residuals from first author



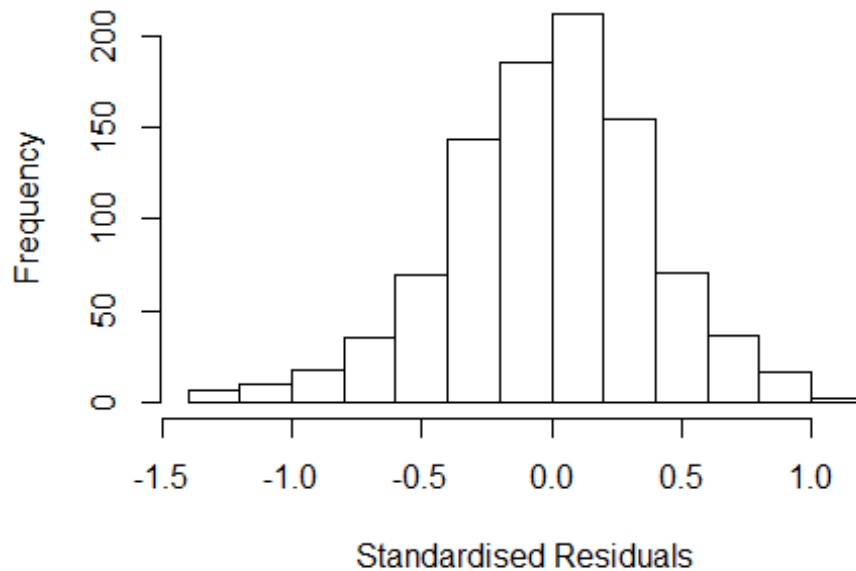
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3096 -0.2575 0.0108 0.2520 1.1611
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14433 0.07508 15.24 <2e-16 ***
## FirstAuthorFemale1 -0.03405 0.02783 -1.22 0.222
## Year1997 0.09993 0.09199 1.09 0.278
## Year1998 0.08864 0.09369 0.95 0.344
## Year1999 0.00558 0.09995 0.06 0.956
## Year2000 0.01494 0.08676 0.17 0.863
## Year2001 0.11546 0.09276 1.24 0.214
## Year2002 0.03391 0.09334 0.36 0.716
## Year2003 -0.00546 0.09092 -0.06 0.952
## Year2004 0.05534 0.08679 0.64 0.524
## Year2005 0.02927 0.09579 0.31 0.760
## Year2006 -0.01323 0.08937 -0.15 0.882
```

```

## Year2007      -0.01746    0.09096   -0.19    0.848
## Year2008      0.14966    0.08647    1.73    0.084 .
## Year2009      0.09636    0.09338    1.03    0.302
## Year2010      0.08943    0.08913    1.00    0.316
## Year2011      0.07095    0.09026    0.79    0.432
## Year2012      0.16528    0.09126    1.81    0.070 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0217, Adjusted R-squared:  0.00409
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 890 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.213  0.874  0.952  0.902  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.076 1          1.037
## Year            1.076 16          1.002

```

## Residuals from last author



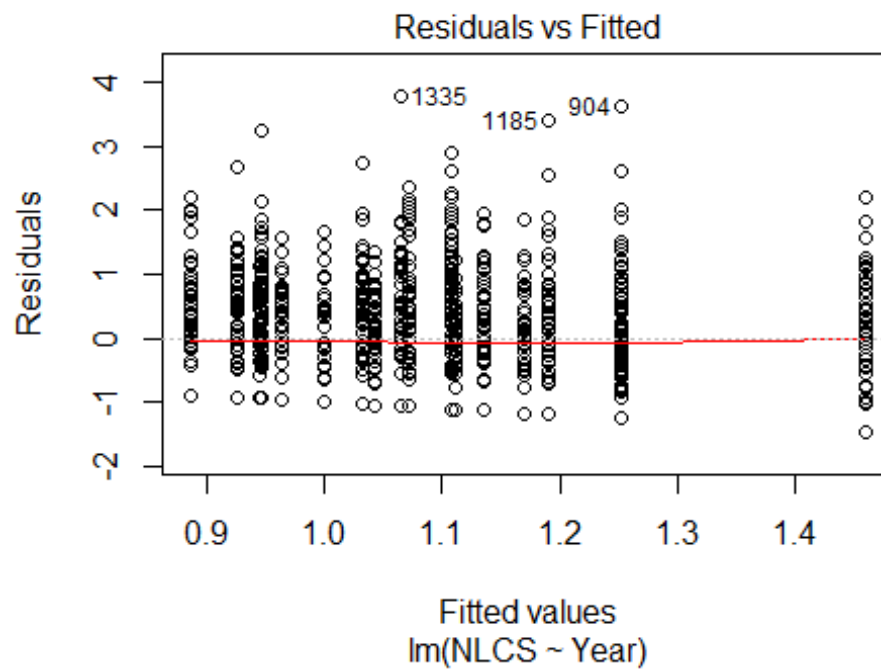
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2991 -0.2545 0.0138 0.2541 1.1672
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14251 0.07446 15.34 <2e-16 ***
## LastAuthorFemale1 -0.04330 0.03579 -1.21 0.227
## Year1997 0.10343 0.09120 1.13 0.257
## Year1998 0.08369 0.09338 0.90 0.370
## Year1999 0.00572 0.10003 0.06 0.954
## Year2000 0.01184 0.08617 0.14 0.891
## Year2001 0.11231 0.09250 1.21 0.225
## Year2002 0.03954 0.09315 0.42 0.671
## Year2003 -0.00968 0.09018 -0.11 0.915
## Year2004 0.05126 0.08601 0.60 0.551
## Year2005 0.02370 0.09537 0.25 0.804
## Year2006 -0.01349 0.08875 -0.15 0.879
```

```

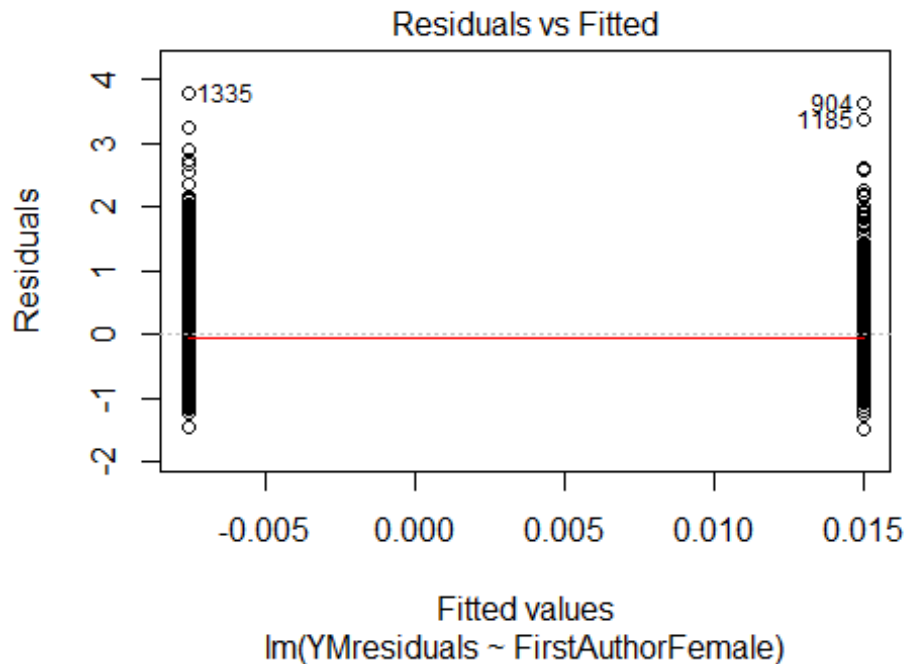
## Year2007          -0.01911      0.09018    -0.21      0.832
## Year2008           0.14369      0.08537      1.68      0.093 .
## Year2009           0.09484      0.09239      1.03      0.305
## Year2010           0.08787      0.08836      0.99      0.320
## Year2011           0.07173      0.08985      0.80      0.425
## Year2012           0.15657      0.08948      1.75      0.081 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.0215, Adjusted R-squared:  0.0038
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 880 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.220  0.873   0.952   0.901   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 960"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1200"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   71   62   79   53  104   64  132  100  117  125  125  129  119  122  127
## 2011 2012
##  146  184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   44   53   38   66   48   79   62   82   88   85   65   65   57   83
## 2011 2012

```

```
## 106 135
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 44 52 37 59 47 79 52 78 84 80 64 62 55 81
## 2011 2012
## 103 133
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 80, df = 16, p-value = 1e-10
```

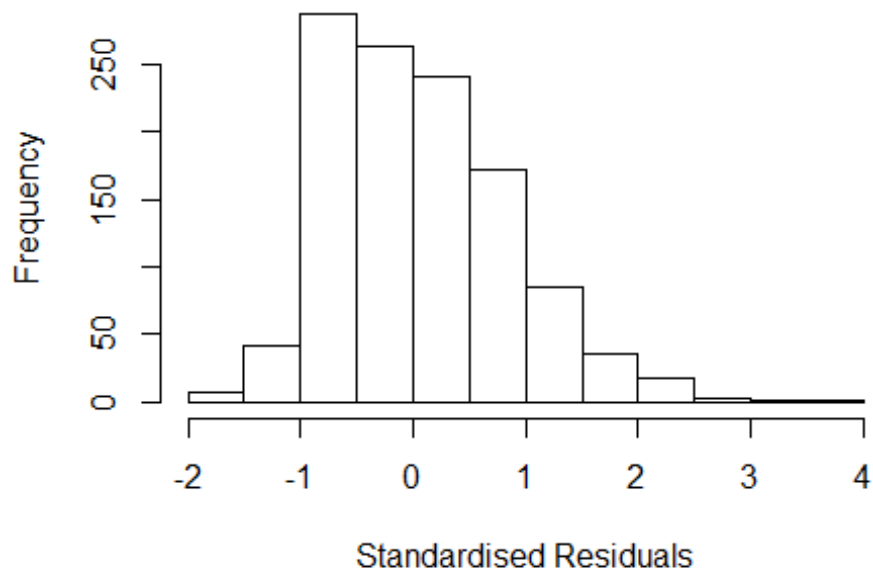


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.19751194217139"
## [1] "Male first author team size 2018 geometric mean: 1.29744078228696"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.17768241766372"
## [1] "Male last author team size 2018 geometric mean: 1.31700719290615"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.448 1          2.109
## LastAuthorFemale  4.457 1          2.111
## UniqueAuthors    1.599 4          1.060
## Year              1.590 16         1.015
```

## Residuals from first and last author and team size



```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 923  17544365628 3.848 2005     1200      2      2.768
## 1181 33947152939 3.747 2007     1200      3      2.745
## 1185 35348964160 4.593 2007     1200      1      3.577
## 1335 42149118387 4.842 2008     1200      1      3.925
## 1794 79951880145 4.181 2011     1200      2      3.410
## 1939 84865215150 3.724 2012     1200      1      2.788
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8594 -0.7598 -0.0372  0.5723  3.9253
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97224    0.09760   9.96 < 2e-16 ***
## FirstAuthorFemale1 0.12417    0.11198   1.11  0.2677
## LastAuthorFemale1 -0.11078    0.11230  -0.99  0.3241
## UniqueAuthors2    0.49511    0.10591   4.68 3.3e-06 ***
## UniqueAuthors3    0.83289    0.11298   7.37 3.2e-13 ***
```



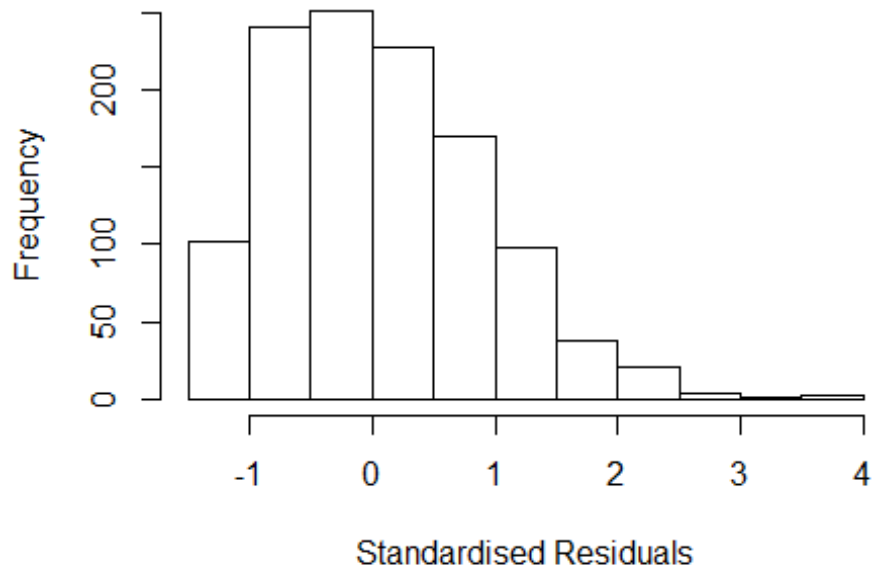
```

## UniqueAuthors4      0.94492      0.28821      3.28      0.0011 **
## UniqueAuthors5      1.22357      0.18994      6.44      1.7e-10 ***
## Year1997             -0.14311      0.14581      -0.98      0.3266
## Year1998             -0.06353      0.13386      -0.47      0.6352
## Year1999             -0.00988      0.13396      -0.07      0.9412
## Year2000             -0.18194      0.12718      -1.43      0.1528
## Year2001             -0.21505      0.11948      -1.80      0.0721 .
## Year2002             -0.12649      0.15015      -0.84      0.3997
## Year2003              0.31085      0.15413      2.02      0.0439 *
## Year2004             -0.05144      0.13685      -0.38      0.7071
## Year2005              0.09393      0.13964      0.67      0.5013
## Year2006             -0.20405      0.13018      -1.57      0.1173
## Year2007              0.02992      0.14481      0.21      0.8364
## Year2008             -0.05553      0.14877      -0.37      0.7090
## Year2009             -0.01454      0.14379      -0.10      0.9195
## Year2010             -0.12892      0.13030      -0.99      0.3227
## Year2011             -0.20166      0.13195      -1.53      0.1267
## Year2012             -0.04957      0.13648      -0.36      0.7165
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.821
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0832
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 767 is an outlier with |weight| = 0 ( < 8.7e-05);
## 86 weights are ~= 1. The remaining 1068 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.018  0.888  0.933  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.66e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 7.338 1          2.709

```

## LastAuthorFemale	7.433	1	2.726
## Year	1.116	16	1.003

### Residuals from first and last author



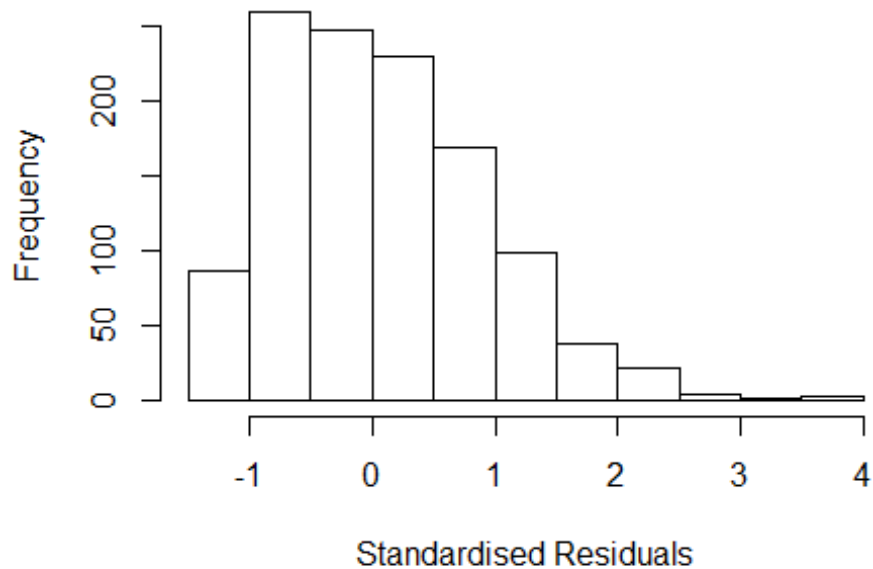
```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 793   1942506038 3.776 2004    1200      2      2.863
## 923   17544365628 3.848 2005    1200      2      2.710
## 1181  33947152939 3.747 2007    1200      3      2.688
## 1185  35348964160 4.593 2007    1200      1      3.518
## 1335  42149118387 4.842 2008    1200      1      3.859
## 1794  79951880145 4.181 2011    1200      2      3.338
## 1939  84865215150 3.724 2012    1200      1      2.707
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3783 -0.6922 -0.0536  0.6109  3.8594
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10324    0.09983   11.05  <2e-16 ***
## FirstAuthorFemale1  0.04591    0.15233    0.30   0.763
## LastAuthorFemale1 -0.02968    0.15185   -0.20   0.845
```

```

## Year1997      -0.12424    0.14371   -0.86    0.387
## Year1998      -0.05777    0.13304   -0.43    0.664
## Year1999       0.00425    0.13554    0.03    0.975
## Year2000      -0.17638    0.12624   -1.40    0.163
## Year2001      -0.15832    0.12564   -1.26    0.208
## Year2002      -0.15331    0.15804   -0.97    0.332
## Year2003       0.25884    0.16277    1.59    0.112
## Year2004      -0.16033    0.13909   -1.15    0.249
## Year2005       0.01845    0.14436    0.13    0.898
## Year2006      -0.30266    0.13541   -2.24    0.026 *
## Year2007      -0.04438    0.15256   -0.29    0.771
## Year2008      -0.12067    0.15874   -0.76    0.447
## Year2009      -0.02735    0.15553   -0.18    0.860
## Year2010      -0.22962    0.13447   -1.71    0.088 .
## Year2011      -0.26018    0.13945   -1.87    0.062 .
## Year2012      -0.10294    0.13956   -0.74    0.461
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.857
## Multiple R-squared:  0.0206, Adjusted R-squared:  0.00504
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 1076 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0056 0.8800 0.9400 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      8.66e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1          1.020
## Year              1.041 16          1.001

```

## Residuals from first author



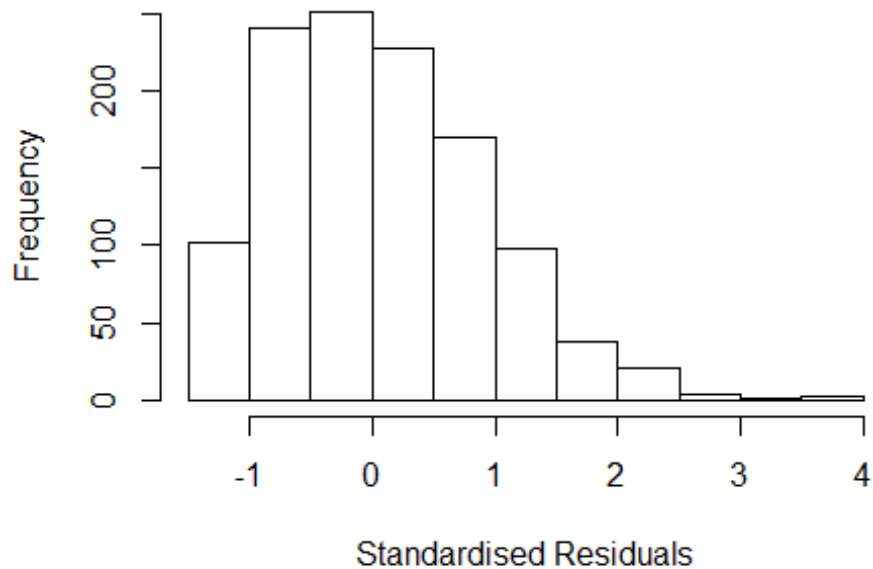
```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 793   1942506038 3.776 2004     1200      2      2.863
## 923   17544365628 3.848 2005     1200      2      2.710
## 1181  33947152939 3.747 2007     1200      3      2.688
## 1185  35348964160 4.593 2007     1200      1      3.518
## 1335  42149118387 4.842 2008     1200      1      3.859
## 1794  79951880145 4.181 2011     1200      2      3.338
## 1939  84865215150 3.724 2012     1200      1      2.707
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3801 -0.6900 -0.0613  0.6121  3.8609
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10103    0.09955   11.06  <2e-16 ***
## FirstAuthorFemale1  0.01933    0.05764    0.34   0.737
## Year1997        -0.12363    0.14370   -0.86   0.390
## Year1998        -0.05720    0.13282   -0.43   0.667
## Year1999         0.00723    0.13544    0.05   0.957
## Year2000        -0.17493    0.12614   -1.39   0.166
```

```

## Year2001      -0.15555    0.12473   -1.25    0.213
## Year2002      -0.15207    0.15807   -0.96    0.336
## Year2003       0.25972    0.16284    1.59    0.111
## Year2004      -0.15763    0.13838   -1.14    0.255
## Year2005       0.02016    0.14425    0.14    0.889
## Year2006      -0.30136    0.13554   -2.22    0.026 *
## Year2007      -0.04244    0.15265   -0.28    0.781
## Year2008      -0.11988    0.15874   -0.76    0.450
## Year2009      -0.02749    0.15536   -0.18    0.860
## Year2010      -0.22785    0.13439   -1.70    0.090 .
## Year2011      -0.25892    0.13951   -1.86    0.064 .
## Year2012      -0.10139    0.13939   -0.73    0.467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.859
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.00582
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 1077 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8800 0.9410 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.66e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.056 1          1.028
## Year            1.056 16          1.002

```

## Residuals from last author



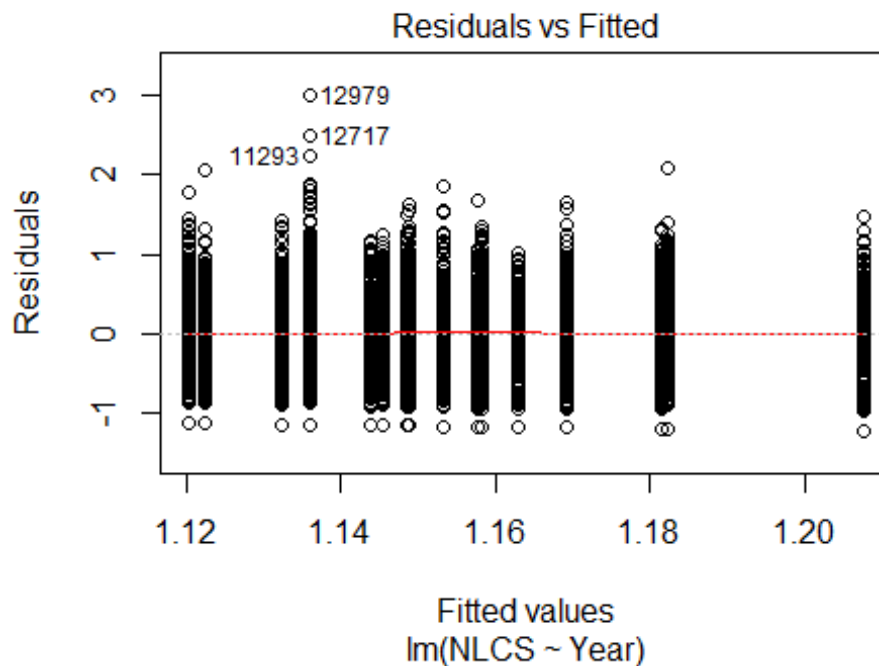
```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 793   1942506038 3.776 2004     1200      2     2.863
## 923   17544365628 3.848 2005     1200      2     2.710
## 1181  33947152939 3.747 2007     1200      3     2.688
## 1185  35348964160 4.593 2007     1200      1     3.518
## 1335  42149118387 4.842 2008     1200      1     3.859
## 1794  79951880145 4.181 2011     1200      2     3.338
## 1939  84865215150 3.724 2012     1200      1     2.707
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3746 -0.6919 -0.0608  0.6093  3.8584
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10287    0.09964   11.07  <2e-16 ***
## LastAuthorFemale1 0.01151    0.05721    0.20   0.841
## Year1997       -0.12384    0.14343   -0.86   0.388
## Year1998       -0.05836    0.13253   -0.44   0.660
## Year1999        0.00776    0.13592    0.06   0.954
## Year2000       -0.17379    0.12624   -1.38   0.169
```

```

## Year2001      -0.15456      0.12482      -1.24      0.216
## Year2002      -0.15084      0.15792      -0.96      0.340
## Year2003       0.26027      0.16289       1.60      0.110
## Year2004      -0.15684      0.13871      -1.13      0.258
## Year2005       0.02041      0.14423       0.14      0.887
## Year2006      -0.30011      0.13555      -2.21      0.027 *
## Year2007      -0.04003      0.15271      -0.26      0.793
## Year2008      -0.11923      0.15862      -0.75      0.452
## Year2009      -0.02690      0.15546      -0.17      0.863
## Year2010      -0.22668      0.13464      -1.68      0.093 .
## Year2011      -0.25807      0.13950      -1.85      0.065 .
## Year2012      -0.09959      0.13933      -0.71      0.475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.86
## Multiple R-squared:  0.0204, Adjusted R-squared:  0.00572
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 1075 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0069 0.8820 0.9400 0.9100 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           8.66e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1155"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1201"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 485 485 477 476 564 615 589 541 562 617 763 905 734 776 879

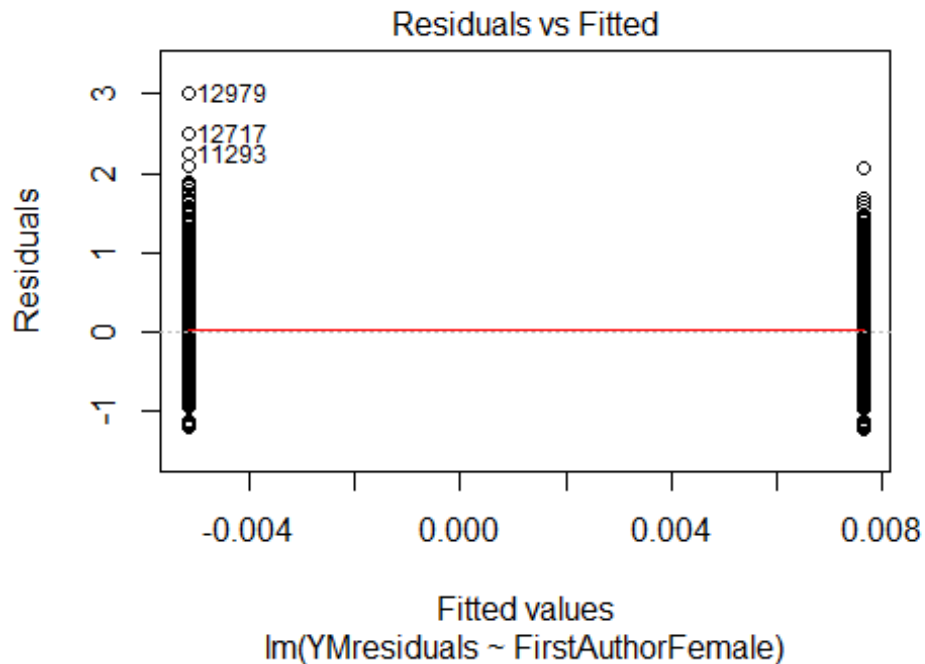
```

```
## 2011 2012
## 948 964
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 357 359 352 366 384 318 449 439 458 499 622 749 589 631 731
## 2011 2012
## 795 801
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 344 333 331 346 360 303 421 414 415 440 563 690 546 573 700
## 2011 2012
## 751 757
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 100, df = 16, p-value = 1e-14
```



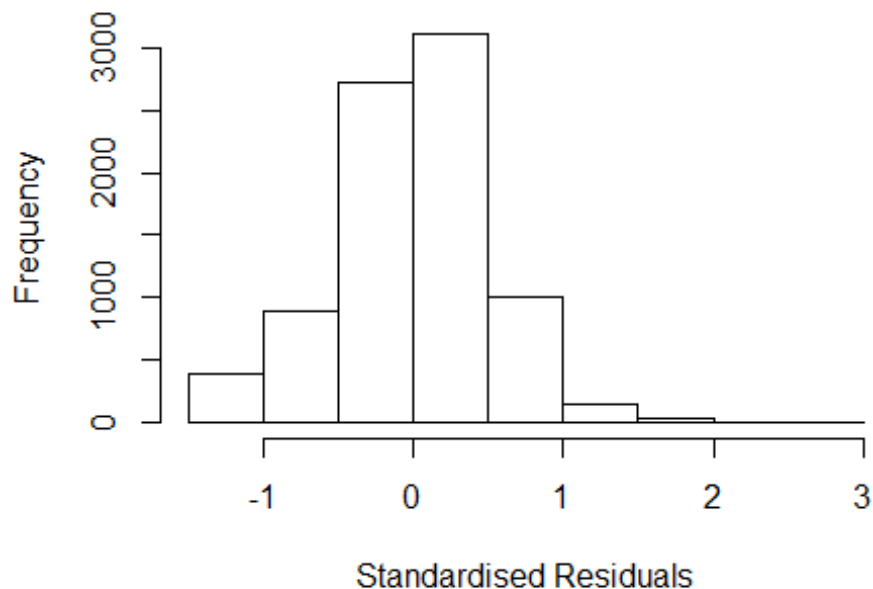
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 7e-04
```





```
## [1] "Female first author team size 2018 geometric mean: 1.75236711643952"
## [1] "Male first author team size 2018 geometric mean: 1.57512879124591"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 76000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.6550687126864"
## [1] "Male last author team size 2018 geometric mean: 1.67399187342747"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 70000, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.289 1      1.135
## LastAuthorFemale  1.303 1      1.141
## UniqueAuthors    1.055 4      1.007
## Year              1.074 16     1.002
```

## Residuals from first and last author and team size



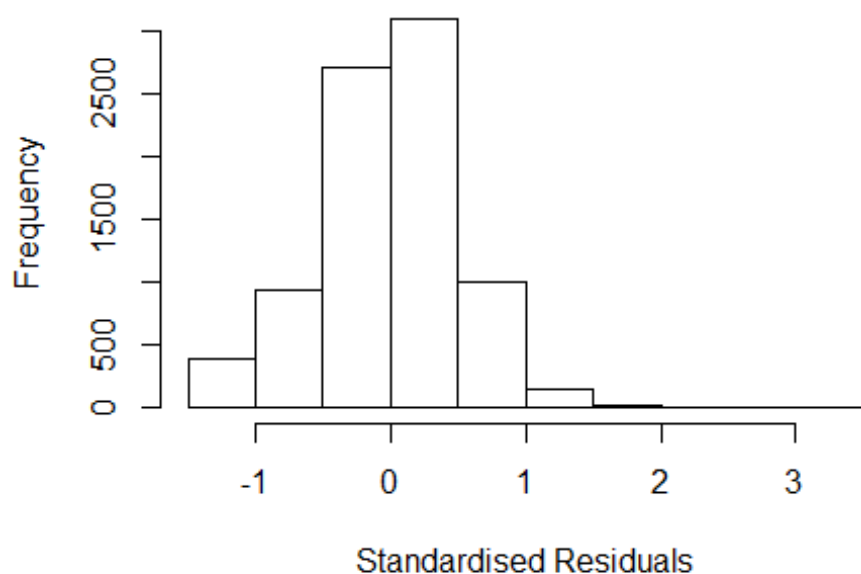
```
## [1] "List of 2 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 12717 84860189956 3.633 2012    1201      3      2.618
## 12979 84857747100 4.135 2012    1201      3      2.942
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3328 -0.3064  0.0181  0.3148  2.9416
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.037298   0.031288   33.15  <2e-16 ***
## FirstAuthorFemale1 0.014506   0.012519    1.16   0.247
## LastAuthorFemale1 0.000603   0.012831    0.05   0.963
## UniqueAuthors2    0.162653   0.013619   11.94  <2e-16 ***
## UniqueAuthors3    0.178069   0.015657   11.37  <2e-16 ***
## UniqueAuthors4    0.184923   0.021227    8.71  <2e-16 ***
## UniqueAuthors5    0.250630   0.021927   11.43  <2e-16 ***
## Year1997          0.018563   0.044358    0.42   0.676
## Year1998          0.036099   0.041890    0.86   0.389
```

```

## Year1999      0.022568    0.041732    0.54    0.589
## Year2000      0.022145    0.041308    0.54    0.592
## Year2001      0.033284    0.041487    0.80    0.422
## Year2002      0.044541    0.039146    1.14    0.255
## Year2003      0.049852    0.038248    1.30    0.192
## Year2004      0.078660    0.038536    2.04    0.041 *
## Year2005      0.029790    0.038721    0.77    0.442
## Year2006     -0.008033    0.037109   -0.22    0.829
## Year2007     -0.000291    0.035195   -0.01    0.993
## Year2008      0.029779    0.036783    0.81    0.418
## Year2009      0.031629    0.036147    0.87    0.382
## Year2010      0.046718    0.035807    1.30    0.192
## Year2011     -0.007520    0.036968   -0.20    0.839
## Year2012     -0.021998    0.037506   -0.59    0.558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.0371, Adjusted R-squared:  0.0346
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(5643,6927,7873,8043)
## are outliers with |weight| = 0 ( < 1.2e-05);
## 677 weights are ~= 1. The remaining 7606 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0015 0.8620 0.9520 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.305 1      1.142
## LastAuthorFemale 1.314 1      1.146
## Year      1.034 16      1.001

```

## Residuals from first and last author



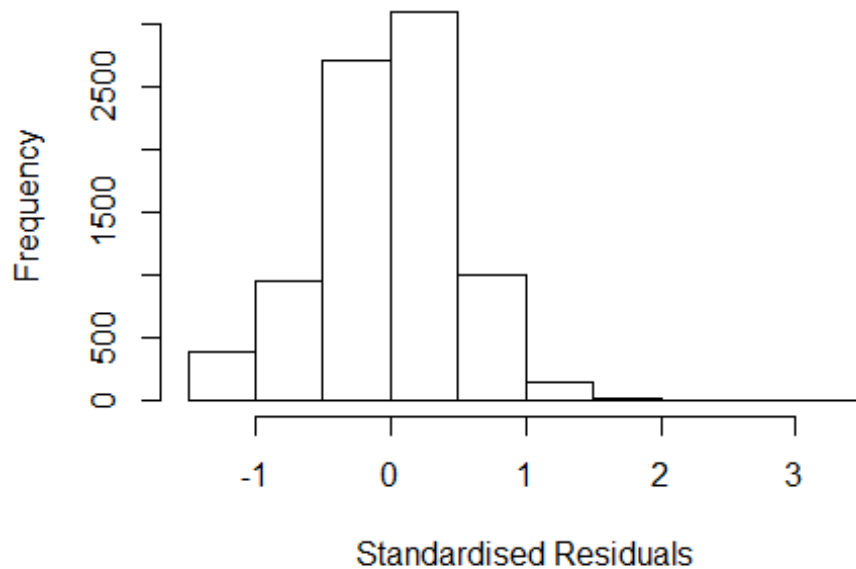
```
## [1] "List of 2 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 12717 84860189956 3.633 2012    1201      3    2.514
## 12979 84857747100 4.135 2012    1201      3    3.016
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2140 -0.3187  0.0173  0.3202  3.0159
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11151    0.03165   35.12  <2e-16 ***
## FirstAuthorFemale1  0.03092    0.01280    2.42   0.016 *
## LastAuthorFemale1 -0.01835    0.01310   -1.40   0.161
## Year1997         0.01545    0.04546    0.34   0.734
## Year1998         0.03960    0.04287    0.92   0.356
## Year1999         0.02694    0.04284    0.63   0.529
## Year2000         0.03289    0.04259    0.77   0.440
## Year2001         0.03675    0.04228    0.87   0.385
## Year2002         0.05539    0.04014    1.38   0.168
## Year2003         0.06232    0.03897    1.60   0.110
## Year2004         0.08991    0.03939    2.28   0.022 *
```

```

## Year2005          0.03849    0.03975    0.97    0.333
## Year2006          0.00789    0.03775    0.21    0.834
## Year2007          0.01075    0.03594    0.30    0.765
## Year2008          0.04382    0.03758    1.17    0.244
## Year2009          0.05679    0.03683    1.54    0.123
## Year2010          0.06615    0.03640    1.82    0.069 .
## Year2011          0.00871    0.03777    0.23    0.818
## Year2012          0.00762    0.03803    0.20    0.841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.00327,    Adjusted R-squared:  0.0011
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(6927,7873,8043)
## are outliers with |weight| = 0 ( < 1.2e-05);
## 685 weights are ~ = 1. The remaining 7599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0086 0.8630 0.9510 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.21e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1          1.009
## Year              1.017 16          1.001

```

## Residuals from first author



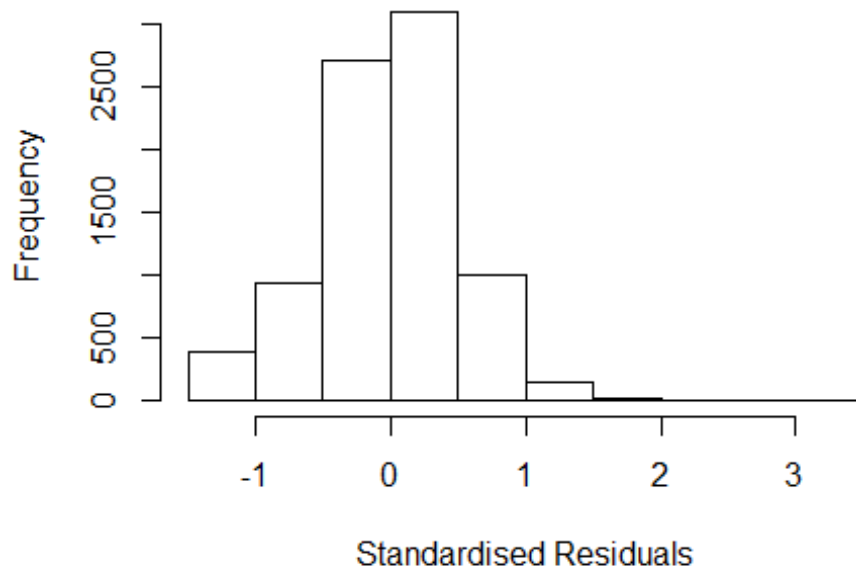
```
## [1] "List of 2 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 12717 84860189956 3.633 2012    1201      3    2.514
## 12979 84857747100 4.135 2012    1201      3    3.016
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2206 -0.3193  0.0177  0.3204  3.0192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10876    0.03160   35.08  <2e-16 ***
## FirstAuthorFemale1 0.02103    0.01135    1.85   0.064 .
## Year1997        0.01583    0.04548    0.35   0.728
## Year1998        0.03907    0.04291    0.91   0.362
## Year1999        0.02756    0.04285    0.64   0.520
## Year2000        0.03200    0.04257    0.75   0.452
## Year2001        0.03666    0.04230    0.87   0.386
## Year2002        0.05543    0.04016    1.38   0.168
## Year2003        0.06199    0.03899    1.59   0.112
## Year2004        0.09077    0.03940    2.30   0.021 *
## Year2005        0.03869    0.03978    0.97   0.331
```

```

## Year2006          0.00816    0.03778    0.22    0.829
## Year2007          0.01053    0.03596    0.29    0.770
## Year2008          0.04381    0.03760    1.17    0.244
## Year2009          0.05697    0.03686    1.55    0.122
## Year2010          0.06648    0.03641    1.83    0.068 .
## Year2011          0.00795    0.03778    0.21    0.833
## Year2012          0.00708    0.03804    0.19    0.852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.00307,    Adjusted R-squared:  0.00103
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(6927,7873,8043)
## are outliers with |weight| = 0 ( < 1.2e-05);
## 700 weights are ~ = 1. The remaining 7584 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0081 0.8620 0.9510 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001

```

## Residuals from last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 12717 84860189956 3.633 2012    1201      3    2.514
## 12979 84857747100 4.135 2012    1201      3    3.016
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2087 -0.3167  0.0173  0.3200  3.0085
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11543    0.03161   35.28  <2e-16 ***
## LastAuthorFemale1 -0.00114    0.01159   -0.10    0.922
## Year1997        0.01777    0.04537    0.39    0.695
## Year1998        0.04068    0.04290    0.95    0.343
## Year1999        0.02838    0.04277    0.66    0.507
## Year2000        0.03392    0.04250    0.80    0.425
## Year2001        0.03933    0.04223    0.93    0.352
## Year2002        0.05737    0.04011    1.43    0.153
## Year2003        0.06409    0.03892    1.65    0.100 .
## Year2004        0.09323    0.03933    2.37    0.018 *
## Year2005        0.04171    0.03966    1.05    0.293
```

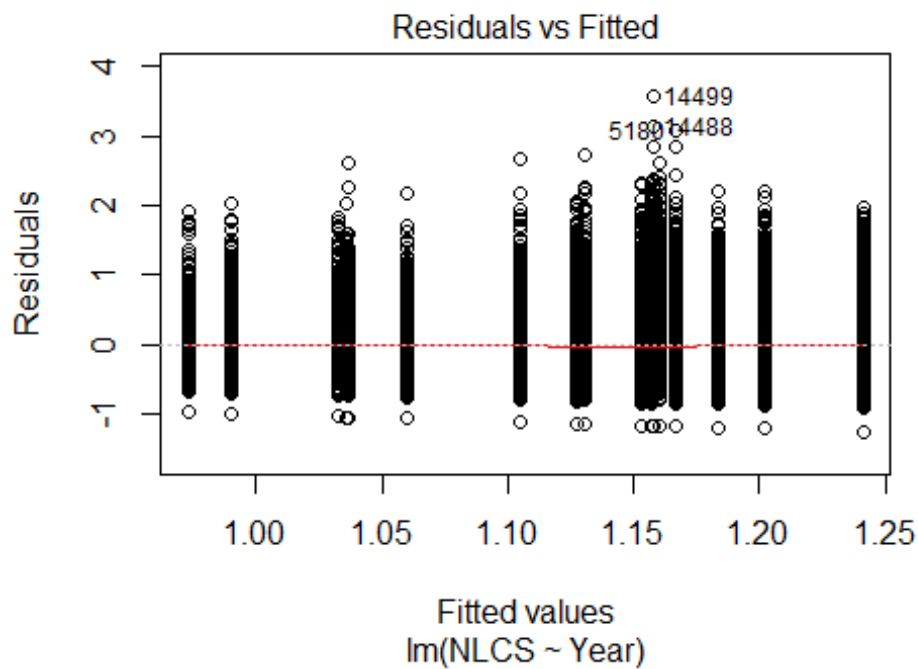


```

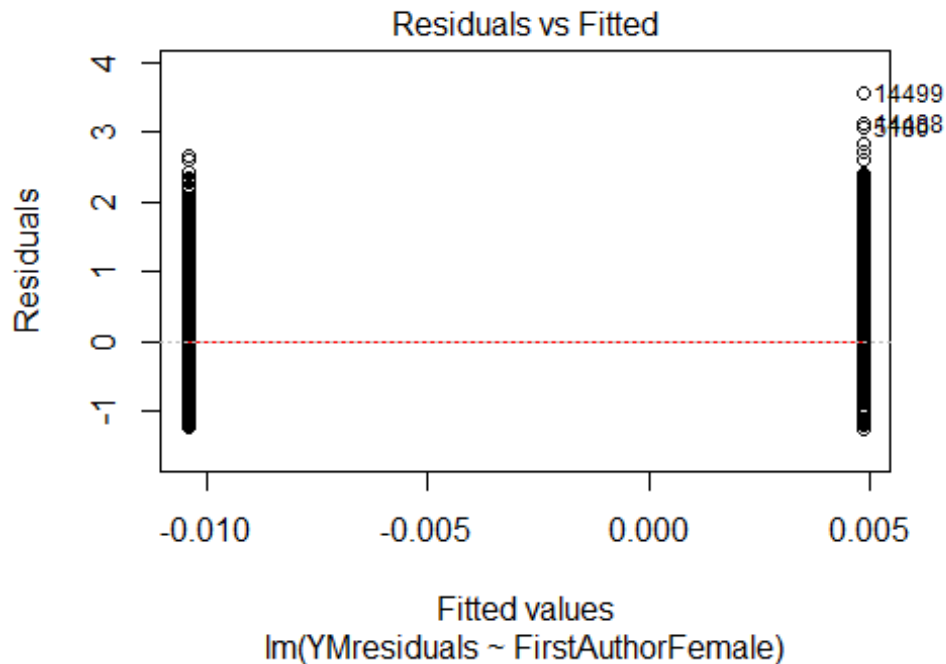
## Year2006      0.01031      0.03772      0.27      0.785
## Year2007      0.01295      0.03587      0.36      0.718
## Year2008      0.04567      0.03751      1.22      0.223
## Year2009      0.05963      0.03677      1.62      0.105
## Year2010      0.06933      0.03629      1.91      0.056 .
## Year2011      0.01141      0.03771      0.30      0.762
## Year2012      0.01106      0.03793      0.29      0.771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.472
## Multiple R-squared:  0.00264,    Adjusted R-squared:  0.00059
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(6927,7873,8043)
## are outliers with |weight| = 0 ( < 1.2e-05);
## 687 weights are ~ = 1. The remaining 7597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0096 0.8620 0.9500 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 8287"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1202"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 478 551 555 586 696 637 752 670 725 747 898 1056 1053 938 1134
## 2011 2012
## 1389 1429
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 377 401 432 468 530 468 604 573 623 628 763 925 882 794 973
## 2011 2012
## 1180 1249
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 373 396 430 462 523 463 601 570 614 623 755 916 873 782 956
## 2011 2012
## 1164 1229
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 150, df = 16, p-value <2e-16
```

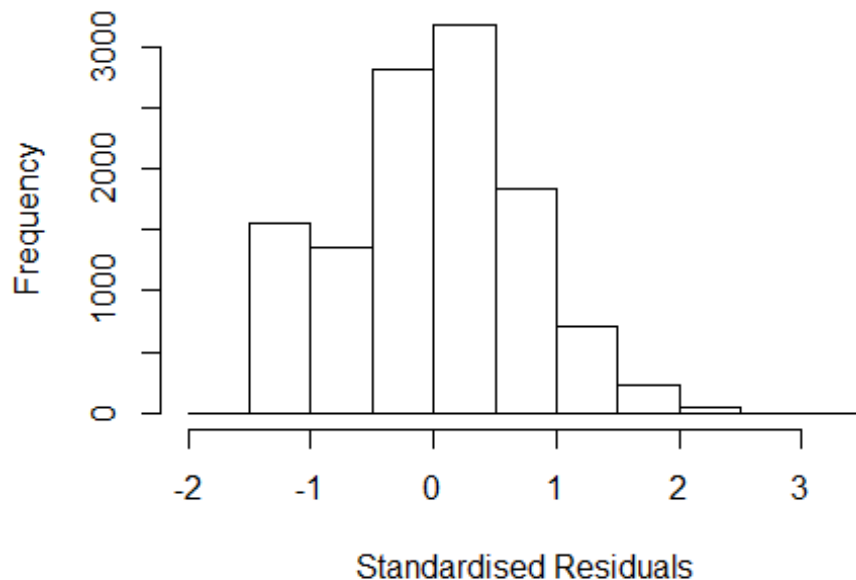


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.1097783403489"
## [1] "Male first author team size 2018 geometric mean: 1.08392031015729"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.09645393285542"
## [1] "Male last author team size 2018 geometric mean: 1.09173810290693"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.090 1          2.468
## LastAuthorFemale  6.080 1          2.466
## UniqueAuthors    1.035 4          1.004
## Year              1.027 16         1.001
```

## Residuals from first and last author and team size



```
## [1] "List of 8 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 3248    0035180284 3.635 2001    1202     1    2.620
## 4240    60949941202 3.850 2002    1202     3    2.768
## 5180    9144258932 4.244 2004    1202     2    3.114
## 5528    85011493186 3.997 2004    1202     1    2.867
## 13194   78751539841 3.763 2011    1202     3    2.670
## 14488   84859894535 4.283 2012    1202     3    2.904
## 14499   84872909005 4.721 2012    1202     4    3.342
## 14642   84857213873 3.996 2012    1202     2    2.903
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7492 -0.4928  0.0194  0.4837  3.3418
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.011358   0.037180   27.20  < 2e-16 ***
## FirstAuthorFemale1 -0.026961   0.036970  -0.73  0.46584
## LastAuthorFemale1  0.010158   0.037028   0.27  0.78384
```

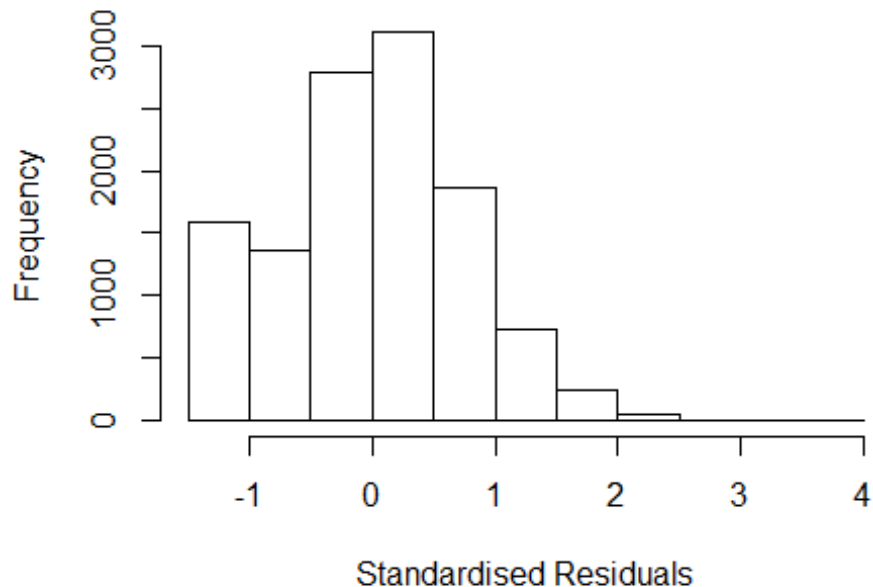
```

## UniqueAuthors2      0.286219    0.026826    10.67 < 2e-16 ***
## UniqueAuthors3      0.305993    0.052603     5.82 6.1e-09 ***
## UniqueAuthors4      0.656751    0.071177     9.23 < 2e-16 ***
## UniqueAuthors5      0.227601    0.126015     1.81 0.07092 .
## Year1997            -0.068897    0.050893    -1.35 0.17584
## Year1998             0.017372    0.048486     0.36 0.72013
## Year1999            -0.060613    0.049549    -1.22 0.22124
## Year2000            -0.000261    0.046751    -0.01 0.99555
## Year2001             0.003724    0.048016     0.08 0.93818
## Year2002             0.070966    0.049568     1.43 0.15226
## Year2003             0.055317    0.047773     1.16 0.24693
## Year2004             0.119069    0.049462     2.41 0.01609 *
## Year2005             0.084969    0.049066     1.73 0.08335 .
## Year2006             0.154994    0.046320     3.35 0.00082 ***
## Year2007             0.095576    0.044748     2.14 0.03271 *
## Year2008             0.199763    0.044096     4.53 6.0e-06 ***
## Year2009             0.138675    0.044673     3.10 0.00191 **
## Year2010             0.071549    0.043647     1.64 0.10118
## Year2011             0.098523    0.043908     2.24 0.02486 *
## Year2012             0.081611    0.044453     1.84 0.06640 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.731
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.024
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 1038 weights are ~= 1. The remaining 10692 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0022 0.8440 0.9500 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.53e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))

```

```
## FirstAuthorFemale 7.203 1 2.684
## LastAuthorFemale 7.191 1 2.682
## Year 1.012 16 1.000
```

### Residuals from first and last author



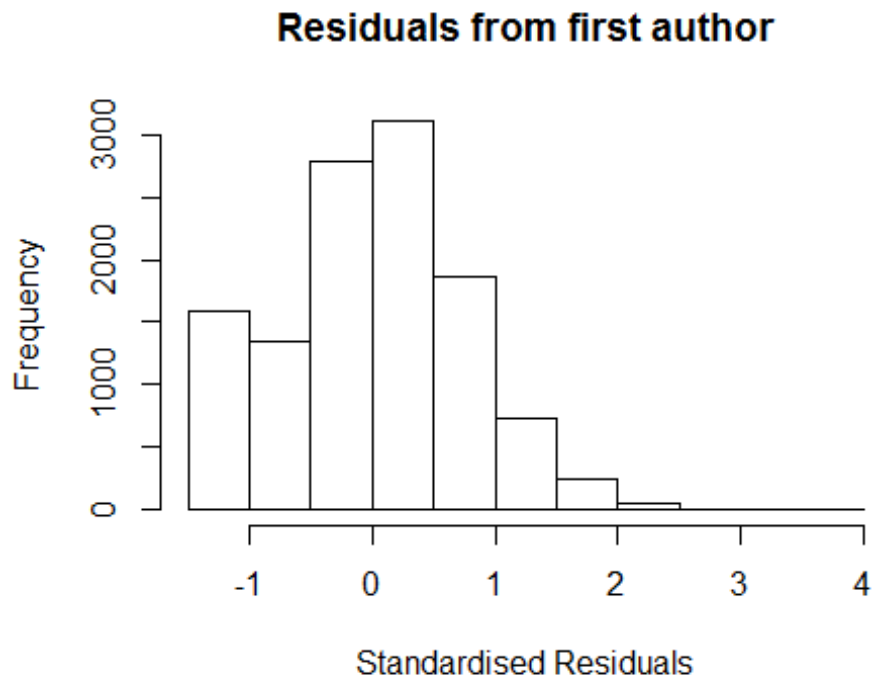
```
## [1] "List of 9 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3248  0035180284 3.635 2001    1202     1    2.599
## 4240  60949941202 3.850 2002    1202     3    2.739
## 4808  61049194206 3.761 2003    1202     2    2.655
## 5180  9144258932 4.244 2004    1202     2    3.085
## 5528  85011493186 3.997 2004    1202     1    2.838
## 13194 78751539841 3.763 2011    1202     3    2.637
## 14488 84859894535 4.283 2012    1202     3    3.158
## 14499 84872909005 4.721 2012    1202     4    3.596
## 14642 84857213873 3.996 2012    1202     2    2.871
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2523 -0.5006  0.0233  0.4933  3.5957
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept)          1.03370    0.03685    28.05 < 2e-16 ***
## FirstAuthorFemale1   0.00873    0.04057     0.22  0.82970
## LastAuthorFemale1   -0.02946    0.04067    -0.72  0.46884
## Year1997             -0.06849    0.05072    -1.35  0.17699
## Year1998              0.03140    0.04846     0.65  0.51701
## Year1999             -0.05618    0.04947    -1.14  0.25606
## Year2000             -0.00209    0.04664    -0.04  0.96425
## Year2001              0.00262    0.04790     0.05  0.95642
## Year2002              0.07764    0.04952     1.57  0.11693
## Year2003              0.06385    0.04749     1.34  0.17882
## Year2004              0.12487    0.04945     2.53  0.01158 *
## Year2005              0.09587    0.04932     1.94  0.05192 .
## Year2006              0.16725    0.04636     3.61  0.00031 ***
## Year2007              0.10344    0.04459     2.32  0.02037 *
## Year2008              0.20984    0.04390     4.78  1.8e-06 ***
## Year2009              0.15232    0.04474     3.40  0.00066 ***
## Year2010              0.08984    0.04350     2.07  0.03890 *
## Year2011              0.11255    0.04371     2.57  0.01004 *
## Year2012              0.09159    0.04444     2.06  0.03932 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.739
## Multiple R-squared:  0.00869,    Adjusted R-squared:  0.00717
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 11368 is an outlier with |weight| = 0 ( < 8.5e-06);
## 1026 weights are ~= 1. The remaining 10703 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.028  0.847  0.949  0.910  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.53e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))

```

```
## FirstAuthorFemale 1.007 1 1.003
## Year 1.007 16 1.000
```



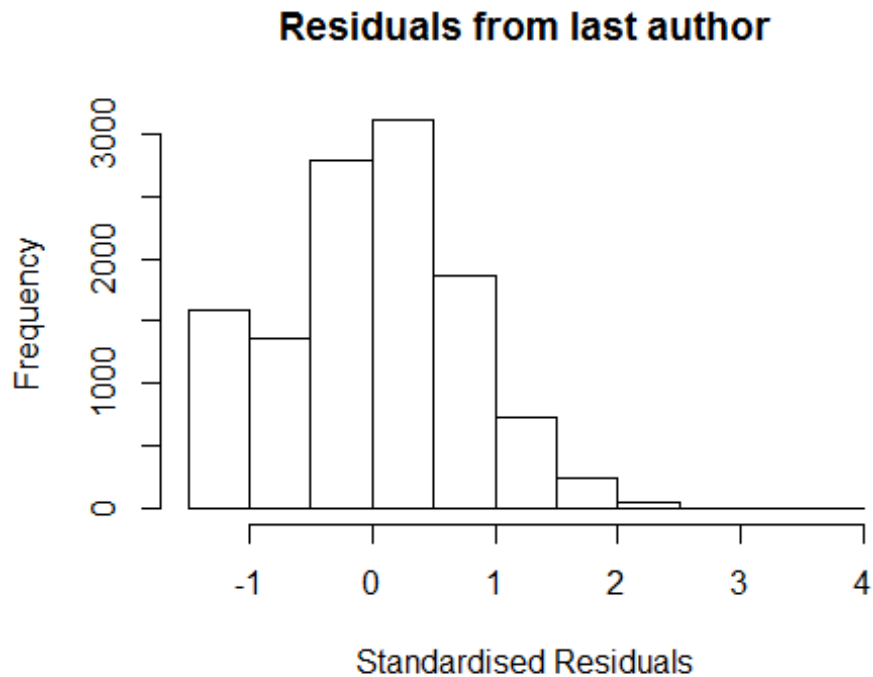
```
## [1] "List of 9 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3248   0035180284 3.635 2001    1202     1    2.599
## 4240   60949941202 3.850 2002    1202     3    2.739
## 4808   61049194206 3.761 2003    1202     2    2.655
## 5180   9144258932 4.244 2004    1202     2    3.085
## 5528   85011493186 3.997 2004    1202     1    2.838
## 13194  78751539841 3.763 2011    1202     3    2.637
## 14488  84859894535 4.283 2012    1202     3    3.158
## 14499  84872909005 4.721 2012    1202     4    3.596
## 14642  84857213873 3.996 2012    1202     2    2.871
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.243 -0.502  0.023  0.494  3.597
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03263    0.03677   28.08 < 2e-16 ***
```



```

## FirstAuthorFemale1 -0.01824    0.01518   -1.20  0.22962
## Year1997           -0.06807    0.05071   -1.34  0.17946
## Year1998            0.03197    0.04845    0.66  0.50936
## Year1999           -0.05596    0.04945   -1.13  0.25783
## Year2000           -0.00180    0.04660   -0.04  0.96926
## Year2001            0.00326    0.04787    0.07  0.94564
## Year2002            0.07773    0.04951    1.57  0.11646
## Year2003            0.06447    0.04746    1.36  0.17436
## Year2004            0.12522    0.04941    2.53  0.01128 *
## Year2005            0.09600    0.04932    1.95  0.05159 .
## Year2006            0.16770    0.04633    3.62  0.00030 ***
## Year2007            0.10387    0.04456    2.33  0.01976 *
## Year2008            0.21061    0.04386    4.80  1.6e-06 ***
## Year2009            0.15280    0.04471    3.42  0.00063 ***
## Year2010            0.09035    0.04346    2.08  0.03767 *
## Year2011            0.11293    0.04370    2.58  0.00977 **
## Year2012            0.09185    0.04442    2.07  0.03866 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.739
## Multiple R-squared:  0.00864,    Adjusted R-squared:  0.0072
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 11368 is an outlier with |weight| = 0 ( < 8.5e-06);
## 1016 weights are ~= 1. The remaining 10713 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0278 0.8470 0.9490 0.9100 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.53e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.005 1          1.003
## Year              1.005 16          1.000

```



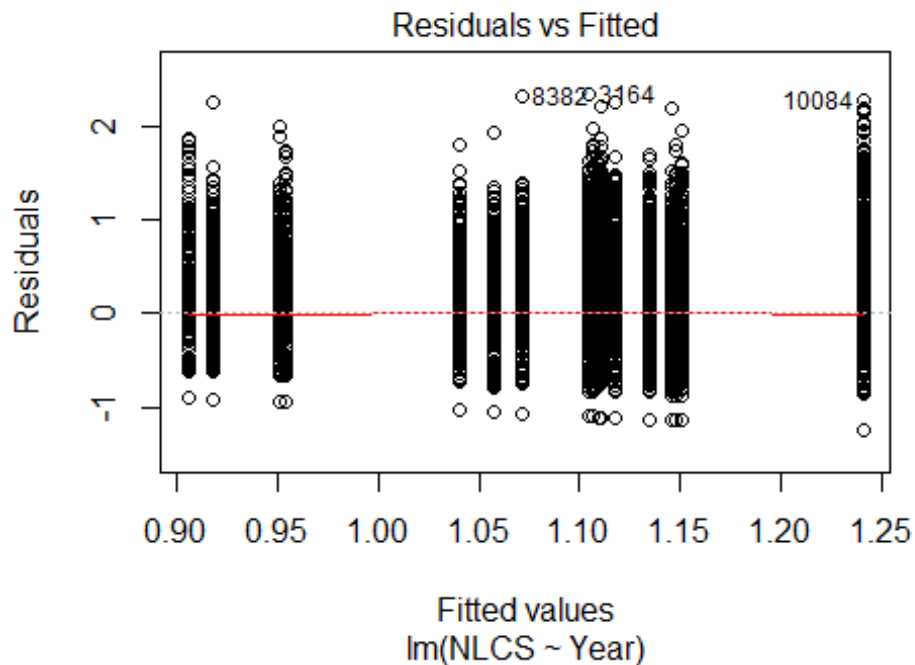
```
## [1] "List of 9 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 3248    0035180284 3.635 2001    1202     1    2.599
## 4240    60949941202 3.850 2002    1202     3    2.739
## 4808    61049194206 3.761 2003    1202     2    2.655
## 5180    9144258932 4.244 2004    1202     2    3.085
## 5528    85011493186 3.997 2004    1202     1    2.838
## 13194   78751539841 3.763 2011    1202     3    2.637
## 14488   84859894535 4.283 2012    1202     3    3.158
## 14499   84872909005 4.721 2012    1202     4    3.596
## 14642   84857213873 3.996 2012    1202     2    2.871
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2439 -0.5003  0.0236  0.4933  3.5954
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03371    0.03685   28.05  < 2e-16 ***
## LastAuthorFemale1 -0.02142    0.01520   -1.41  0.15898
## Year1997        -0.06829    0.05070   -1.35  0.17805
## Year1998         0.03157    0.04844    0.65  0.51461
```

```

## Year1999      -0.05602      0.04945      -1.13      0.25724
## Year2000      -0.00190      0.04661      -0.04      0.96748
## Year2001       0.00291      0.04786       0.06      0.95145
## Year2002       0.07783      0.04950       1.57      0.11591
## Year2003       0.06417      0.04745       1.35      0.17629
## Year2004       0.12507      0.04941       2.53      0.01138 *
## Year2005       0.09603      0.04932       1.95      0.05152 .
## Year2006       0.16751      0.04632       3.62      0.00030 ***
## Year2007       0.10370      0.04454       2.33      0.01993 *
## Year2008       0.21020      0.04385       4.79      1.7e-06 ***
## Year2009       0.15260      0.04469       3.41      0.00064 ***
## Year2010       0.09014      0.04344       2.07      0.03803 *
## Year2011       0.11282      0.04369       2.58      0.00983 **
## Year2012       0.09187      0.04439       2.07      0.03853 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.739
## Multiple R-squared:  0.00869,    Adjusted R-squared:  0.00725
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 11368 is an outlier with |weight| = 0 ( < 8.5e-06);
## 1045 weights are ~= 1. The remaining 10684 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0281 0.8470 0.9490 0.9100 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           8.53e-06           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 11730"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1203"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"

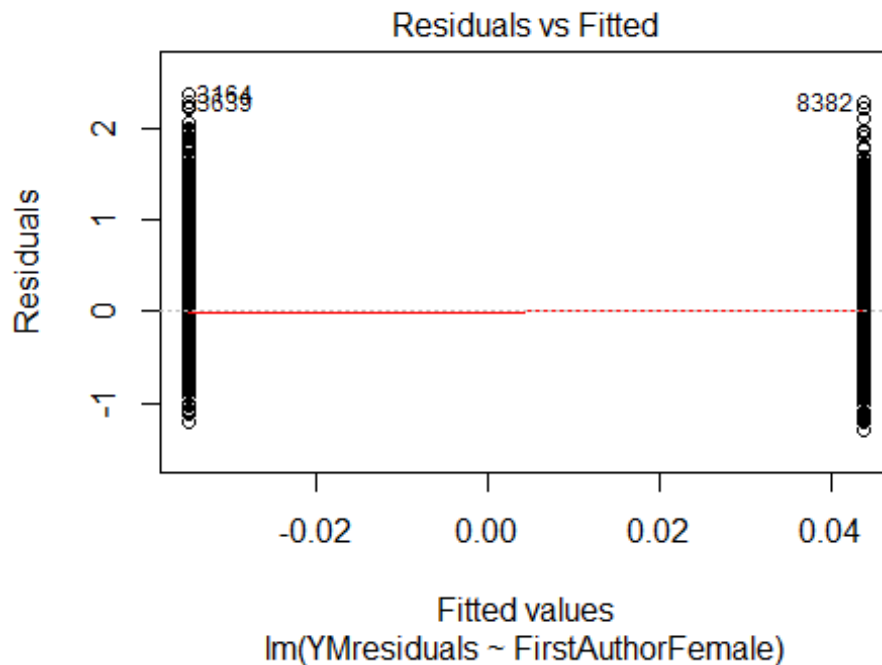
```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 436 494 428 406 454 426 441 443 527 585 600 678 632 572 568
## 2011 2012
## 681 791
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 301 320 373 337 361 327 388 377 454 494 502 580 542 482 488
## 2011 2012
## 585 676
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 287 313 355 323 354 314 370 361 430 479 488 557 517 464 466
## 2011 2012
## 558 641
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 90, df = 16, p-value = 2e-12
```



```
##
## Bartlett test of homogeneity of variances
##
```

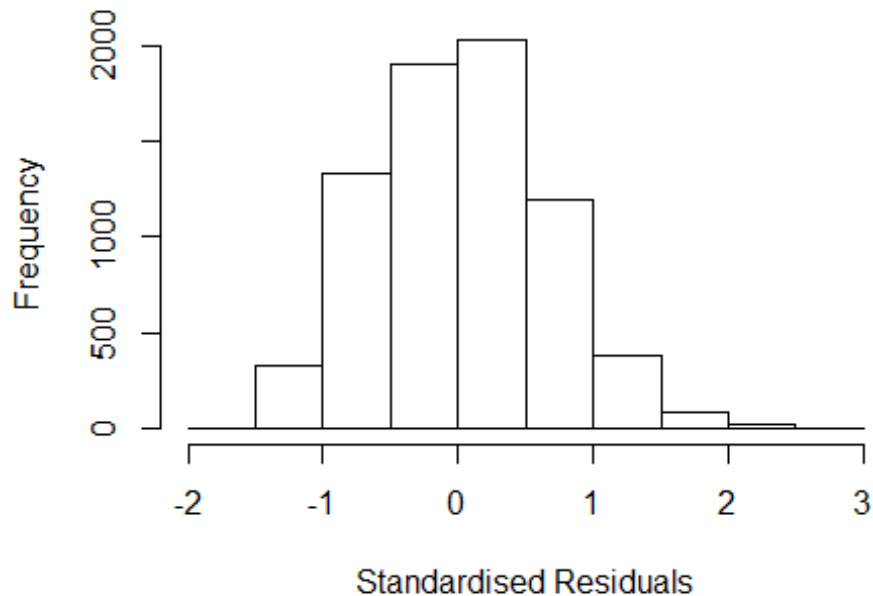
```
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 18, df = 1, p-value = 3e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 1.6161559593756"
## [1] "Male first author team size 2018 geometric mean: 1.31372647228538"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 2e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.56212192988035"
## [1] "Male last author team size 2018 geometric mean: 1.37372923578093"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 36000, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.714 1          1.309
## LastAuthorFemale  1.702 1          1.305
```

## UniqueAuthors	1.097	4	1.012
## Year	1.103	16	1.003

## Residuals from first and last author and team size



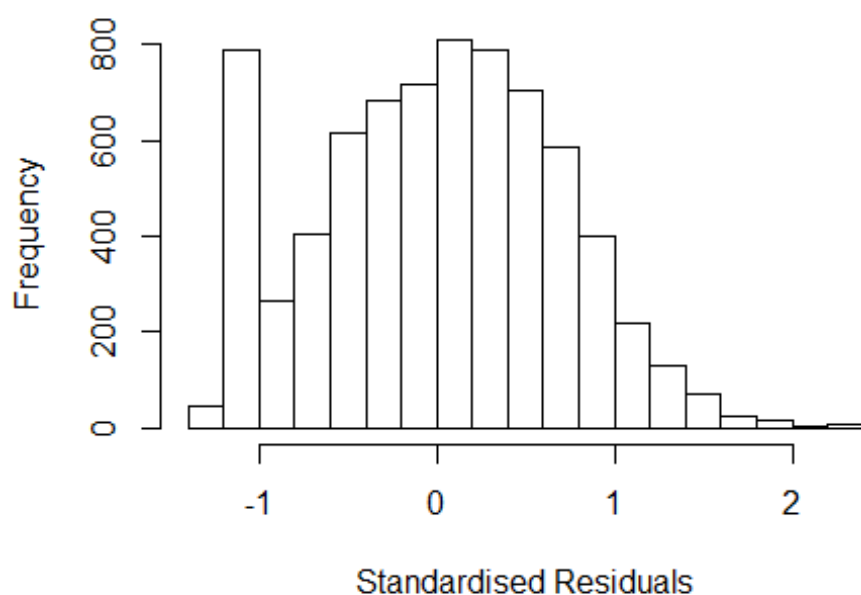
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3164 0036755089 3.44 2002    1203      3      2.519
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5467 -0.4615  0.0215  0.4612  2.5192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7496    0.0421   17.79 < 2e-16 ***
## FirstAuthorFemale1 0.0300    0.0209    1.44  0.14984
## LastAuthorFemale1 0.0218    0.0209    1.04  0.29671
## UniqueAuthors2    0.3757    0.0209   17.99 < 2e-16 ***
## UniqueAuthors3    0.4602    0.0247   18.63 < 2e-16 ***
## UniqueAuthors4    0.5092    0.0343   14.86 < 2e-16 ***
## UniqueAuthors5    0.5866    0.0451   13.00 < 2e-16 ***
```

```

## Year1997          0.0576      0.0578      1.00  0.31941
## Year1998         -0.0135      0.0551     -0.25  0.80615
## Year1999          0.0666      0.0557      1.20  0.23138
## Year2000          0.1458      0.0531      2.75  0.00606 **
## Year2001          0.1435      0.0541      2.65  0.00797 **
## Year2002          0.1712      0.0552      3.10  0.00192 **
## Year2003          0.2008      0.0568      3.54  0.00041 ***
## Year2004          0.2240      0.0518      4.32  1.5e-05 ***
## Year2005          0.2354      0.0515      4.57  4.8e-06 ***
## Year2006          0.2480      0.0506      4.90  9.9e-07 ***
## Year2007          0.2143      0.0506      4.24  2.3e-05 ***
## Year2008          0.1759      0.0511      3.44  0.00058 ***
## Year2009          0.1593      0.0520      3.06  0.00221 **
## Year2010          0.1146      0.0501      2.29  0.02219 *
## Year2011          0.1694      0.0518      3.27  0.00107 **
## Year2012          0.2851      0.0541      5.27  1.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.0979, Adjusted R-squared:  0.0952
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 617 weights are ~= 1. The remaining 6660 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.146  0.860  0.952  0.914  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.37e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.994 1      1.412
## LastAuthorFemale  1.990 1      1.411
## Year              1.019 16      1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2541 -0.5093  0.0285  0.4934  2.3944
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.84230    0.04316   19.51 < 2e-16 ***
## FirstAuthorFemale1 0.07659    0.02378    3.22  0.00128 **
## LastAuthorFemale1  0.00709    0.02391    0.30  0.76682
## Year1997         0.05826    0.05841    1.00  0.31862
## Year1998         0.01119    0.05669    0.20  0.84347
## Year1999         0.06718    0.05892    1.14  0.25425
## Year2000         0.16610    0.05459    3.04  0.00235 **
## Year2001         0.15829    0.05473    2.89  0.00384 **
## Year2002         0.20332    0.05758    3.53  0.00042 ***
## Year2003         0.22356    0.05927    3.77  0.00016 ***
## Year2004         0.24430    0.05393    4.53  6.0e-06 ***
## Year2005         0.26499    0.05317    4.98  6.4e-07 ***
```

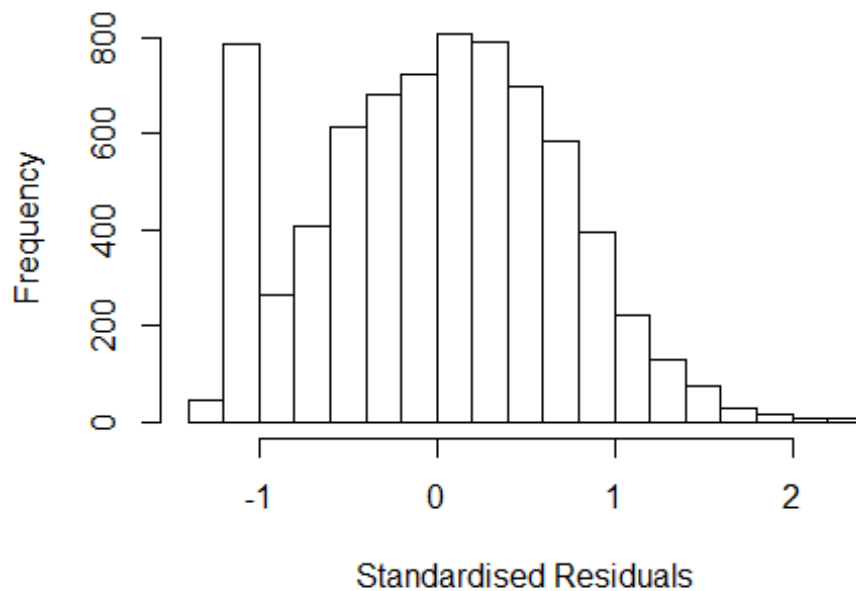


```

## Year2006          0.25761      0.05226      4.93  8.4e-07 ***
## Year2007          0.24284      0.05237      4.64  3.6e-06 ***
## Year2008          0.20435      0.05282      3.87  0.00011 ***
## Year2009          0.23700      0.05419      4.37  1.2e-05 ***
## Year2010          0.18072      0.05252      3.44  0.00058 ***
## Year2011          0.21017      0.05351      3.93  8.7e-05 ***
## Year2012          0.32808      0.05627      5.83  5.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.723
## Multiple R-squared:  0.0188, Adjusted R-squared:  0.0164
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 575 weights are ~= 1. The remaining 6702 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.251  0.867  0.949  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.37e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.011 1      1.005
## Year              1.011 16      1.000

```

## Residuals from first author



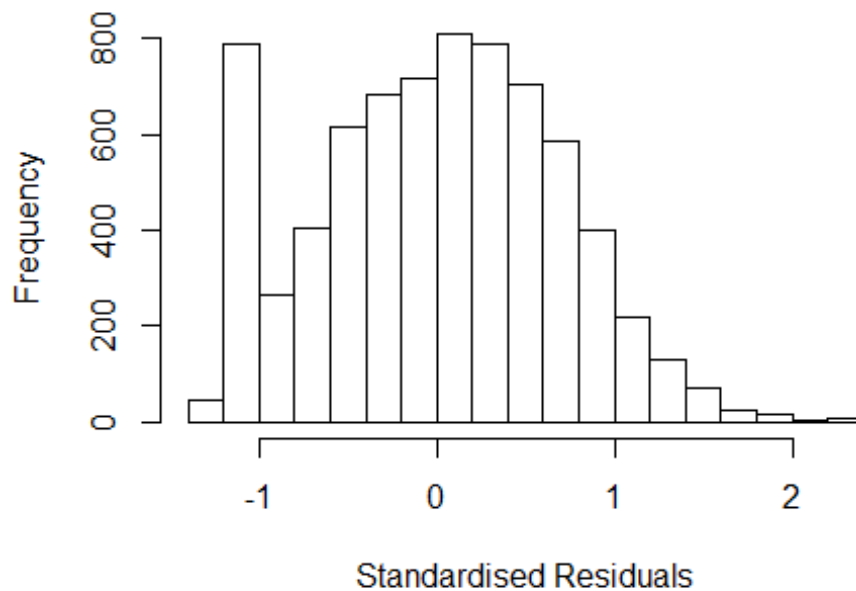
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2530 -0.5100 0.0286 0.4923 2.3936
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8430 0.0431 19.57 < 2e-16 ***
## FirstAuthorFemale1 0.0816 0.0169 4.82 1.4e-06 ***
## Year1997 0.0583 0.0584 1.00 0.31814
## Year1998 0.0111 0.0567 0.20 0.84458
## Year1999 0.0671 0.0589 1.14 0.25444
## Year2000 0.1660 0.0546 3.04 0.00236 **
## Year2001 0.1584 0.0547 2.89 0.00382 **
## Year2002 0.2034 0.0576 3.53 0.00041 ***
## Year2003 0.2240 0.0592 3.78 0.00016 ***
## Year2004 0.2445 0.0539 4.54 5.8e-06 ***
## Year2005 0.2652 0.0532 4.99 6.2e-07 ***
## Year2006 0.2578 0.0522 4.93 8.3e-07 ***
```

```

## Year2007          0.2427      0.0524      4.64  3.6e-06 ***
## Year2008          0.2045      0.0528      3.87  0.00011 ***
## Year2009          0.2371      0.0542      4.38  1.2e-05 ***
## Year2010          0.1811      0.0525      3.45  0.00057 ***
## Year2011          0.2102      0.0535      3.93  8.6e-05 ***
## Year2012          0.3284      0.0563      5.84  5.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.723
## Multiple R-squared:  0.0188, Adjusted R-squared:  0.0165
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 574 weights are ~= 1. The remaining 6703 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.251  0.867  0.949  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.37e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.009 1      1.004
## Year              1.009 16      1.000

```

## Residuals from last author



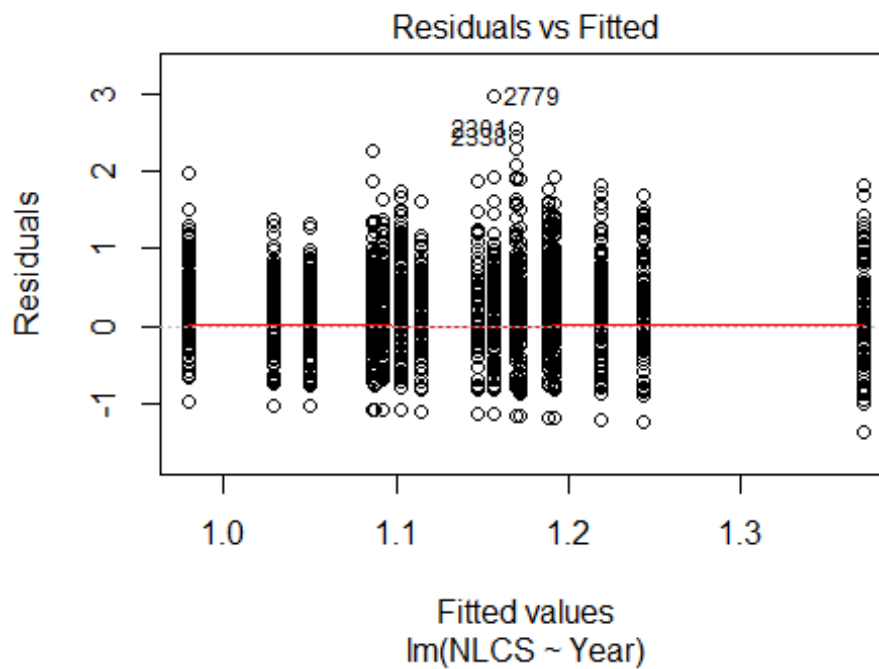
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2429 -0.5049  0.0311  0.4943  2.3832
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8516    0.0433   19.68 < 2e-16 ***
## LastAuthorFemale1 0.0620    0.0170    3.65 0.00027 ***
## Year1997        0.0564    0.0584    0.96 0.33462
## Year1998        0.0122    0.0567    0.21 0.83018
## Year1999        0.0663    0.0592    1.12 0.26263
## Year2000        0.1689    0.0547    3.09 0.00204 **
## Year2001        0.1591    0.0547    2.91 0.00364 **
## Year2002        0.2052    0.0577    3.56 0.00038 ***
## Year2003        0.2219    0.0594    3.74 0.00019 ***
## Year2004        0.2441    0.0541    4.51 6.6e-06 ***
## Year2005        0.2656    0.0533    4.98 6.6e-07 ***
## Year2006        0.2604    0.0524    4.97 7.0e-07 ***
```

```

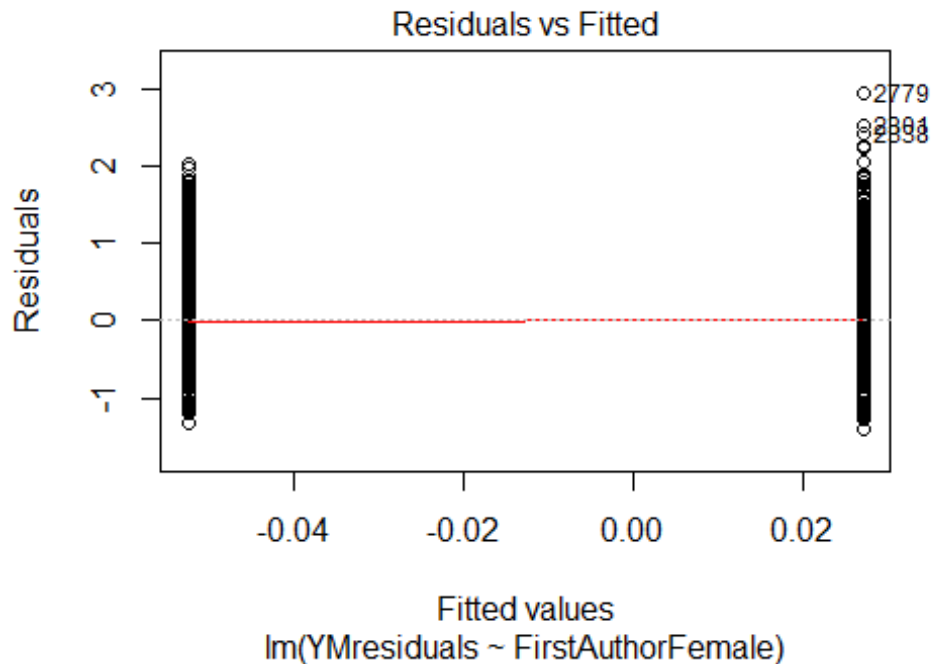
## Year2007          0.2470      0.0526      4.70  2.7e-06 ***
## Year2008          0.2060      0.0530      3.89  0.00010 ***
## Year2009          0.2389      0.0543      4.40  1.1e-05 ***
## Year2010          0.1808      0.0527      3.43  0.00060 ***
## Year2011          0.2133      0.0537      3.97  7.1e-05 ***
## Year2012          0.3293      0.0564      5.84  5.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.723
## Multiple R-squared:  0.0174, Adjusted R-squared:  0.0151
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 592 weights are ~= 1. The remaining 6685 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.866  0.949  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.37e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7277"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1204"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 128 151 163 148 165 132 171 176 193 174 206 252 239 209 247
## 2011 2012
## 247 299
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 98 113 103 105 78 137 130 151 135 163 194 190 162 182
## 2011 2012

```

```
## 197 244
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 80 95 109 101 100 75 131 128 141 127 150 186 184 149 173
## 2011 2012
## 190 237
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```

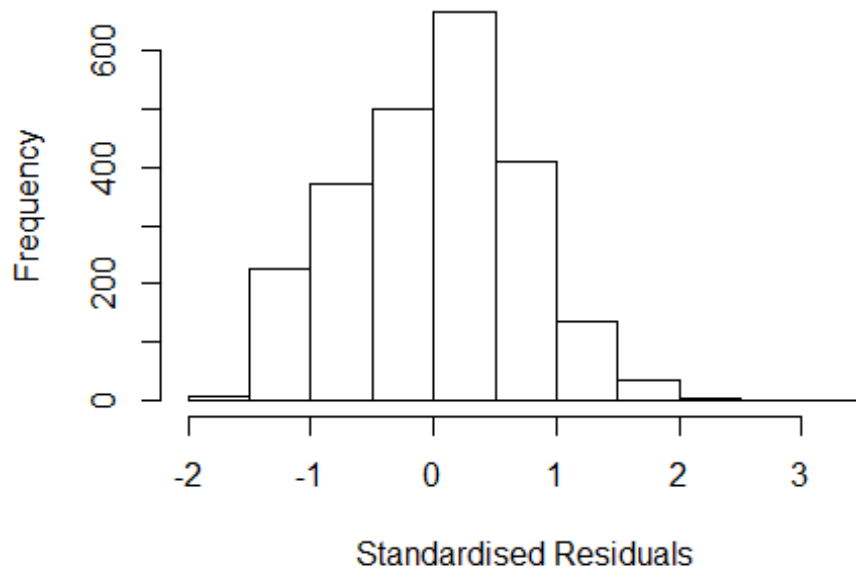


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.71429442210909"
## [1] "Male first author team size 2018 geometric mean: 1.42731889764969"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7500, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.4902197998388"
## [1] "Male last author team size 2018 geometric mean: 1.55093394439056"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.865 1      1.366
## LastAuthorFemale  1.938 1      1.392
## UniqueAuthors     1.200 4      1.023
## Year               1.226 16     1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2301 33947152939 3.747 2007    1200     3     2.678
## 2779 79954549466 4.139 2009    1204     2     3.142
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5857 -0.5154  0.0463  0.4987  3.1416
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13162    0.08324   13.59 < 2e-16 ***
## FirstAuthorFemale1 0.00315    0.04416    0.07  0.9432
## LastAuthorFemale1 -0.12659    0.04649   -2.72  0.0065 **
## UniqueAuthors2    0.31033    0.04332    7.16 1.1e-12 ***
## UniqueAuthors3    0.42643    0.05252    8.12 7.5e-16 ***
## UniqueAuthors4    0.45901    0.06122    7.50 9.1e-14 ***
## UniqueAuthors5    0.40688    0.05940    6.85 9.4e-12 ***
## Year1997        -0.08500    0.10652   -0.80  0.4250
## Year1998        -0.14034    0.10254   -1.37  0.1713
```

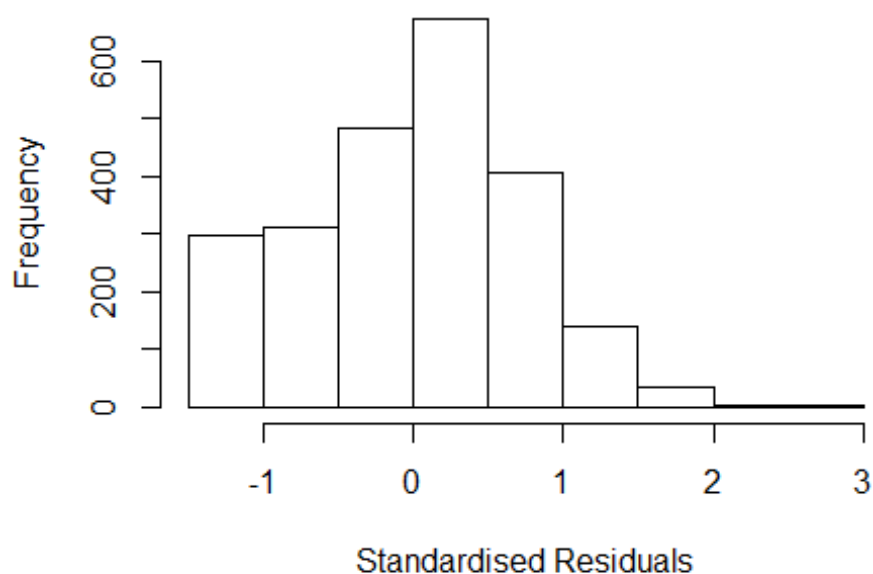


```

## Year1999      -0.10457    0.10348   -1.01    0.3124
## Year2000      -0.18609    0.10652   -1.75    0.0808 .
## Year2001      -0.05657    0.11403   -0.50    0.6199
## Year2002       0.04764    0.11007    0.43    0.6652
## Year2003      -0.11687    0.10232   -1.14    0.2535
## Year2004      -0.12066    0.09961   -1.21    0.2259
## Year2005       0.14377    0.10275    1.40    0.1619
## Year2006       0.00288    0.10249    0.03    0.9776
## Year2007      -0.06310    0.10350   -0.61    0.5422
## Year2008      -0.12456    0.10029   -1.24    0.2144
## Year2009      -0.13423    0.10021   -1.34    0.1806
## Year2010      -0.27982    0.09637   -2.90    0.0037 **
## Year2011      -0.03384    0.09853   -0.34    0.7313
## Year2012      -0.02388    0.09749   -0.24    0.8065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.727
## Multiple R-squared:  0.079, Adjusted R-squared:  0.0703
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 230 weights are ~= 1. The remaining 2126 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0225 0.8630 0.9460 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.24e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.953 1          1.397
## LastAuthorFemale 1.990 1          1.411
## Year              1.077 16          1.002

```

## Residuals from first and last author



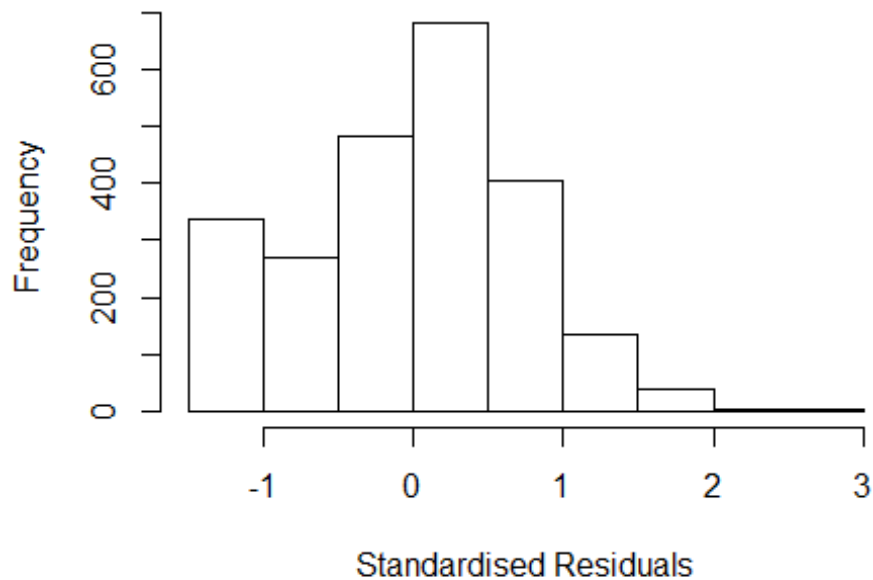
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2301 33947152939 3.747 2007    1200     3    2.573
## 2338 34047099783 3.640 2007    1204     2    2.684
## 2779 79954549466 4.139 2009    1204     2    2.998
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4064 -0.5327  0.0563  0.4977  2.9981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.235289   0.078529   15.73 < 2e-16 ***
## FirstAuthorFemale1  0.066101   0.046727    1.41  0.157
## LastAuthorFemale1 -0.217844   0.048868   -4.46 8.7e-06 ***
## Year1997        -0.050698   0.102920   -0.49  0.622
## Year1998        -0.142158   0.098963   -1.44  0.151
## Year1999        -0.107013   0.098806   -1.08  0.279
## Year2000        -0.156194   0.102595   -1.52  0.128
## Year2001        -0.041718   0.112018   -0.37  0.710
## Year2002         0.062218   0.109331    0.57  0.569
## Year2003        -0.091495   0.101318   -0.90  0.367
```

```

## Year2004          -0.122066    0.096435    -1.27    0.206
## Year2005          0.171078    0.100532     1.70    0.089 .
## Year2006          0.030780    0.102285     0.30    0.763
## Year2007          -0.061623    0.100648    -0.61    0.540
## Year2008          -0.116921    0.099836    -1.17    0.242
## Year2009          -0.094403    0.098261    -0.96    0.337
## Year2010          -0.235194    0.095487    -2.46    0.014 *
## Year2011          -0.000298    0.095616     0.00    0.998
## Year2012          -0.009628    0.095911    -0.10    0.920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.736
## Multiple R-squared:  0.0261, Adjusted R-squared:  0.0186
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 195 weights are ~= 1. The remaining 2161 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0599 0.8510 0.9470 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## Year              1.022 16      1.001

```

## Residuals from first author



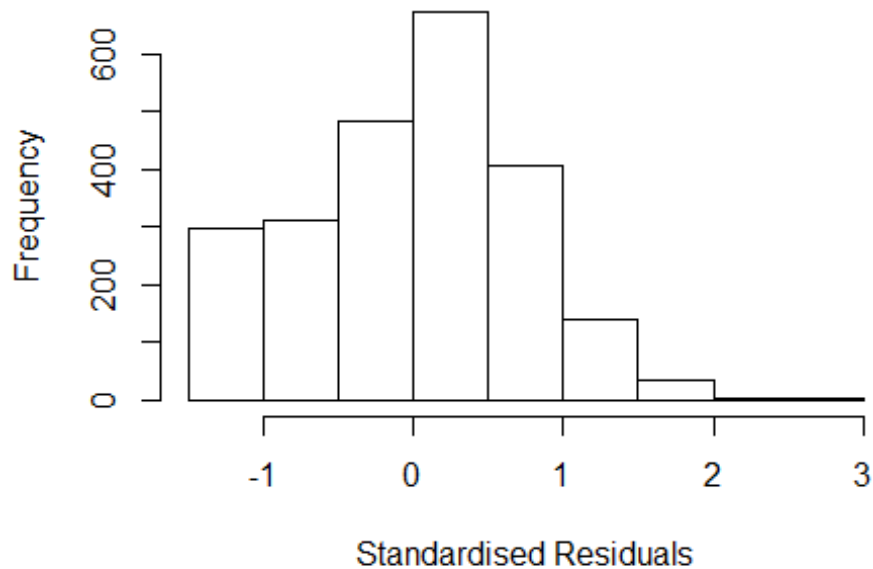
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2301 33947152939 3.747 2007    1200     3    2.573
## 2338 34047099783 3.640 2007    1204     2    2.684
## 2779 79954549466 4.139 2009    1204     2    2.998
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3863 -0.5298  0.0606  0.4926  2.9976
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1982    0.0787   15.22  <2e-16 ***
## FirstAuthorFemale1 -0.0848    0.0344   -2.47   0.014 *
## Year1997         -0.0372    0.1031   -0.36   0.718
## Year1998         -0.1098    0.0995   -1.10   0.270
## Year1999         -0.0804    0.0991   -0.81   0.418
## Year2000         -0.1377    0.1025   -1.34   0.179
## Year2001         -0.0053    0.1131   -0.05   0.963
## Year2002          0.0900    0.1107    0.81   0.416
## Year2003         -0.0807    0.1017   -0.79   0.428
## Year2004         -0.0911    0.0969   -0.94   0.348
```

```

## Year2005          0.1880      0.1017      1.85      0.065 .
## Year2006          0.0515      0.1032      0.50      0.618
## Year2007         -0.0467      0.1011     -0.46      0.644
## Year2008         -0.0975      0.1012     -0.96      0.335
## Year2009         -0.0568      0.0989     -0.57      0.566
## Year2010         -0.2003      0.0960     -2.09      0.037 *
## Year2011          0.0237      0.0960      0.25      0.805
## Year2012          0.0061      0.0962      0.06      0.949
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.737
## Multiple R-squared:  0.0174, Adjusted R-squared:  0.0103
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 201 weights are ~= 1. The remaining 2155 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0608 0.8410 0.9480 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.038 1      1.019
## Year      1.038 16      1.001

```

## Residuals from last author



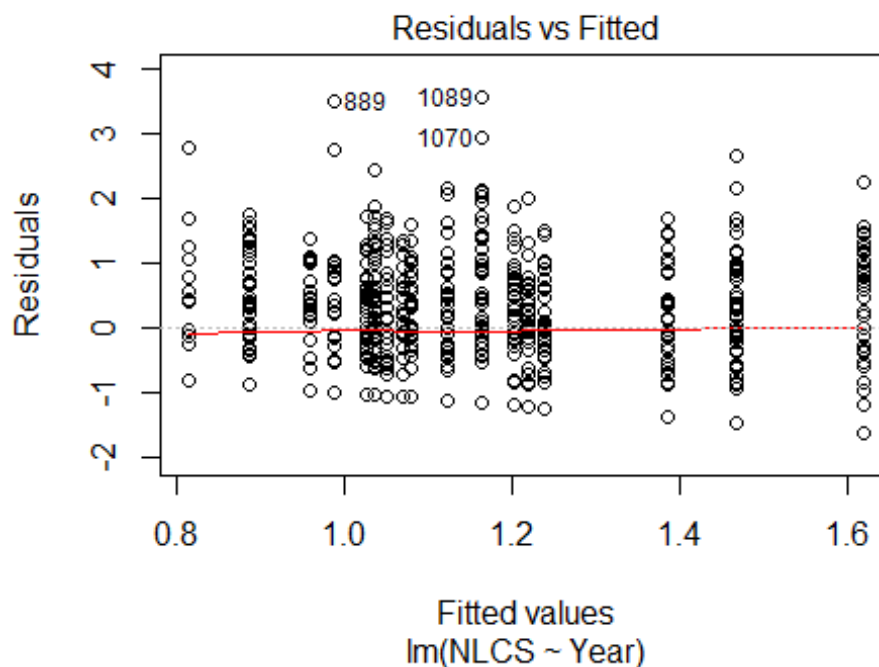
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2301 33947152939 3.747 2007    1200     3    2.573
## 2338 34047099783 3.640 2007    1204     2    2.684
## 2779 79954549466 4.139 2009    1204     2    2.998
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4109 -0.5325  0.0583  0.4995  2.9867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23646    0.07828   15.80 < 2e-16 ***
## LastAuthorFemale1 -0.16876    0.03519   -4.80 1.7e-06 ***
## Year1997        -0.04580    0.10242   -0.45  0.655
## Year1998        -0.13335    0.09820   -1.36  0.175
## Year1999        -0.10162    0.09839   -1.03  0.302
## Year2000        -0.15194    0.10211   -1.49  0.137
## Year2001        -0.03210    0.11130   -0.29  0.773
## Year2002         0.06776    0.10929    0.62  0.535
## Year2003        -0.08765    0.10086   -0.87  0.385
## Year2004        -0.11473    0.09597   -1.20  0.232
```

```

## Year2005          0.17442      0.10047      1.74      0.083 .
## Year2006          0.03638      0.10197      0.36      0.721
## Year2007         -0.05662      0.10043     -0.56      0.573
## Year2008         -0.11049      0.09964     -1.11      0.268
## Year2009         -0.08416      0.09782     -0.86      0.390
## Year2010         -0.22593      0.09487     -2.38      0.017 *
## Year2011          0.00891      0.09507      0.09      0.925
## Year2012         -0.00357      0.09544     -0.04      0.970
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.736
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.0182
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 204 weights are ~= 1. The remaining 2152 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0623 0.8520 0.9470 0.9100 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2356"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1205"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   65   74   64   49   61   55   59   54   63   67   69   90   78   41   32
## 2011 2012
##   70   96
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 41 36 34 34 35 33 41 43 52 46 51 74 65 30 24
## 2011 2012
## 57 86
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 35 34 34 35 33 41 43 52 46 51 74 65 29 23
## 2011 2012
## 57 86
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 37, df = 16, p-value = 0.002
```

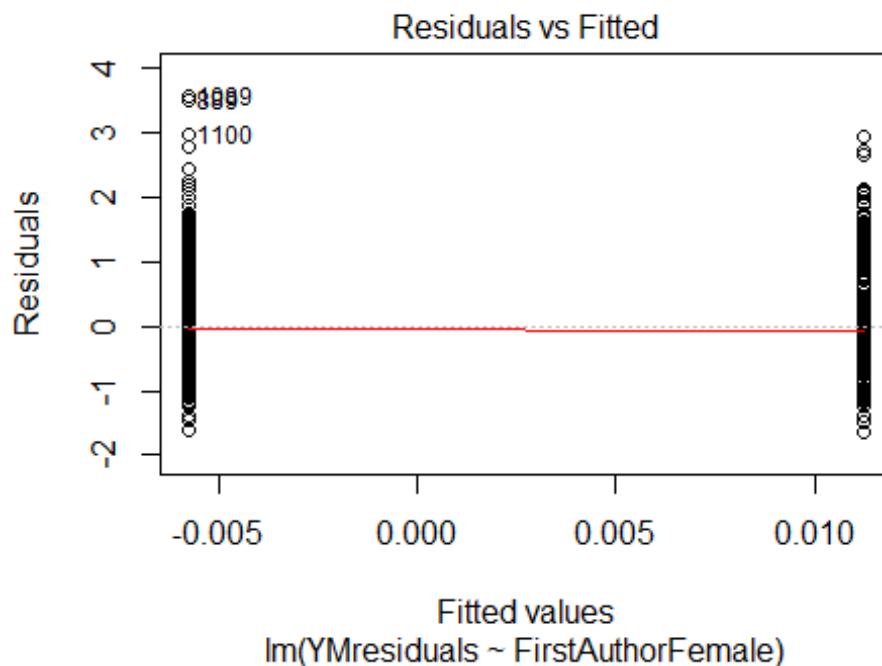


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.47, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 1"
## [1] "Male first author team size 2018 geometric mean: 1.03925922603184"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



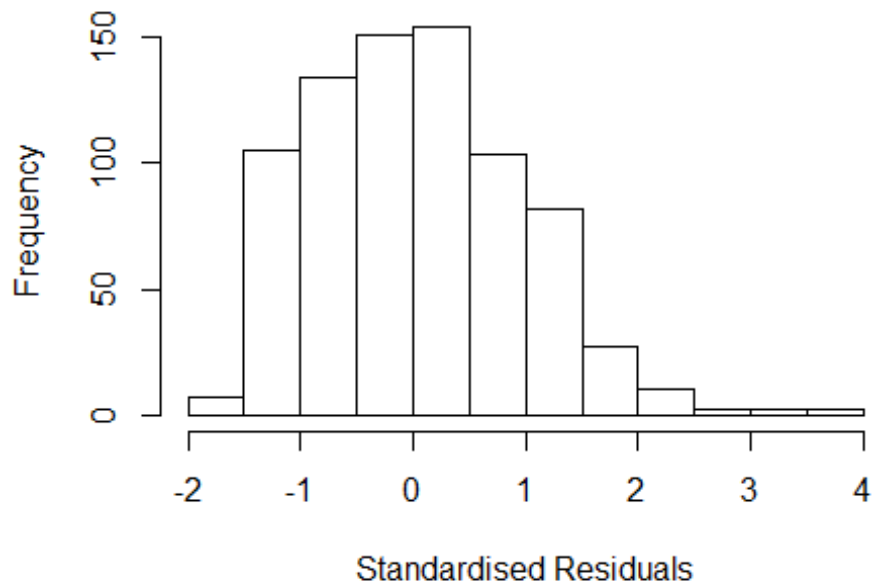
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 68, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.08005973889231"
## [1] "Male last author team size 2018 geometric mean: 1"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 85, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 129.79 1      11.392
## LastAuthorFemale  126.15 1      11.232
## UniqueAuthors    209.45 4       1.950
## Year              51.42 16       1.131
```

## Residuals from first and last author and team size



```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 877  70450198432 3.724 2009    1205     1    2.962
## 889  69949103890 4.473 2009    1205     1    3.728
## 992  79958058988 4.134 2011    1205     1    2.694
## 1070 84871070627 4.115 2012    1205     1    3.097
## 1089 84872909005 4.721 2012    1202     4    3.531
## 1100 84873811421 4.115 2012    1205     1    3.114
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6381 -0.7162 -0.0169  0.6226  3.7285
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.2058     0.1391   8.67 < 2e-16 ***
## FirstAuthorFemale1  0.1488     0.4372   0.34  0.734
## LastAuthorFemale1 -0.1311     0.4403  -0.30  0.766
## UniqueAuthors2    0.1892     0.3161   0.60  0.550
## UniqueAuthors3   -0.6144     0.4630  -1.33  0.185
```

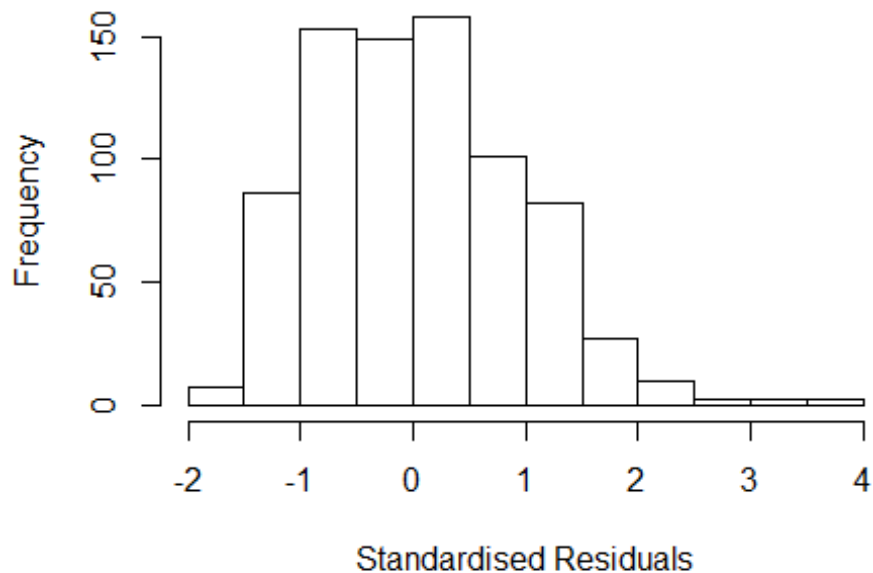
```

## UniqueAuthors4      0.0441      0.2586      0.17      0.865
## UniqueAuthors5      1.1080      0.1527      7.26 9.9e-13 ***
## Year1997             0.0268      0.1879      0.14      0.887
## Year1998             0.0121      0.1894      0.06      0.949
## Year1999            -0.2895      0.1981     -1.46      0.144
## Year2000            -0.2009      0.1998     -1.01      0.315
## Year2001            -0.1516      0.1976     -0.77      0.443
## Year2002             0.4146      0.2374      1.75      0.081 .
## Year2003            -0.1445      0.1955     -0.74      0.460
## Year2004            -0.1440      0.1758     -0.82      0.413
## Year2005             0.1624      0.1951      0.83      0.406
## Year2006            -0.1908      0.1723     -1.11      0.268
## Year2007            -0.2169      0.1719     -1.26      0.207
## Year2008            -0.3694      0.1815     -2.03      0.042 *
## Year2009            -0.4613      0.2168     -2.13      0.034 *
## Year2010            -0.5815      0.2347     -2.48      0.013 *
## Year2011             0.2167      0.2004      1.08      0.280
## Year2012            -0.2053      0.1898     -1.08      0.280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.928
## Multiple R-squared:  0.0607, Adjusted R-squared:  0.0334
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 703 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0697 0.8870 0.9420 0.9130 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.28e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 27.491 1          5.243

```

```
## LastAuthorFemale 27.357 1 5.230
## Year 1.166 16 1.005
```

### Residuals from first and last author



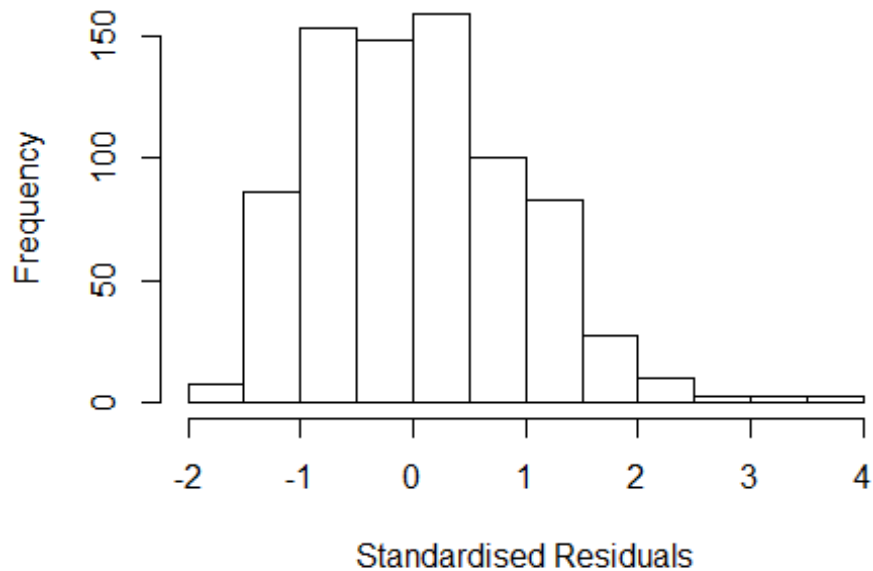
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 877  70450198432 3.724 2009    1205      1    2.960
## 889  69949103890 4.473 2009    1205      1    3.733
## 992  79958058988 4.134 2011    1205      1    2.685
## 1070 84871070627 4.115 2012    1205      1    3.094
## 1089 84872909005 4.721 2012    1202      4    3.725
## 1100 84873811421 4.115 2012    1205      1    3.119
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6429 -0.7158 -0.0195  0.6217  3.7335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.1904     0.1417   8.40  <2e-16 ***
## FirstAuthorFemale1  0.1726     0.3658   0.47   0.637
## LastAuthorFemale1 -0.1480     0.3657  -0.40   0.686
## Year1997         0.0426     0.1890   0.23   0.822
```

```

## Year1998          0.0115      0.1905      0.06      0.952
## Year1999          -0.2413     0.2006     -1.20     0.229
## Year2000          -0.1829     0.2013     -0.91     0.364
## Year2001          -0.1281     0.2002     -0.64     0.523
## Year2002           0.4279     0.2390      1.79     0.074 .
## Year2003          -0.1268     0.1971     -0.64     0.520
## Year2004          -0.1233     0.1775     -0.70     0.487
## Year2005           0.1848     0.1953      0.95     0.344
## Year2006          -0.1773     0.1744     -1.02     0.310
## Year2007          -0.2027     0.1745     -1.16     0.246
## Year2008          -0.3577     0.1838     -1.95     0.052 .
## Year2009          -0.4509     0.2185     -2.06     0.039 *
## Year2010          -0.5602     0.2344     -2.39     0.017 *
## Year2011           0.2344     0.2015      1.16     0.245
## Year2012          -0.1943     0.1920     -1.01     0.312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.925
## Multiple R-squared:  0.0572, Adjusted R-squared:  0.0349
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 707 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0663 0.8880 0.9430 0.9130 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.28e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1          1.033
## Year              1.066 16          1.002

```

## Residuals from first author



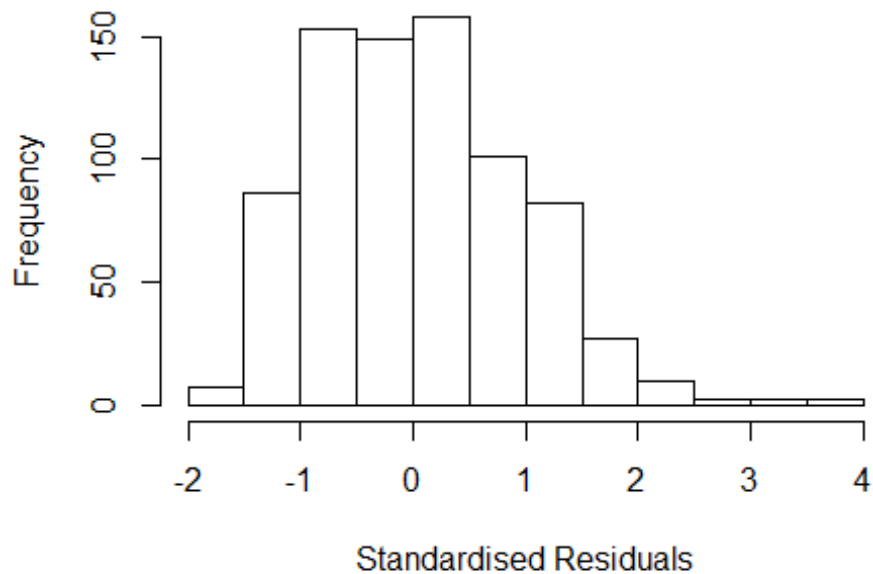
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 877  70450198432 3.724 2009    1205      1    2.960
## 889  69949103890 4.473 2009    1205      1    3.733
## 992  79958058988 4.134 2011    1205      1    2.685
## 1070 84871070627 4.115 2012    1205      1    3.094
## 1089 84872909005 4.721 2012    1202      4    3.725
## 1100 84873811421 4.115 2012    1205      1    3.119
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6448 -0.7153 -0.0196  0.6192  3.7352
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1898     0.1417   8.40 2.2e-16 ***
## FirstAuthorFemale1  0.0274     0.0721   0.38  0.704
## Year1997          0.0375     0.1880   0.20  0.842
## Year1998          0.0114     0.1906   0.06  0.952
## Year1999         -0.2415     0.2006  -1.20  0.229
## Year2000         -0.1788     0.2005  -0.89  0.373
## Year2001         -0.1285     0.2003  -0.64  0.521
```

```

## Year2002          0.4276      0.2390      1.79      0.074 .
## Year2003         -0.1230      0.1966     -0.63      0.532
## Year2004         -0.1235      0.1775     -0.70      0.487
## Year2005          0.1777      0.1953      0.91      0.363
## Year2006         -0.1776      0.1745     -1.02      0.309
## Year2007         -0.2033      0.1744     -1.17      0.244
## Year2008         -0.3586      0.1839     -1.95      0.052 .
## Year2009         -0.4521      0.2186     -2.07      0.039 *
## Year2010         -0.5607      0.2345     -2.39      0.017 *
## Year2011          0.2341      0.2016      1.16      0.246
## Year2012         -0.1934      0.1922     -1.01      0.315
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.925
## Multiple R-squared:  0.057, Adjusted R-squared:  0.0359
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 704 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0657 0.8860 0.9400 0.9130 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.28e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.063 1          1.031
## Year            1.063 16          1.002

```

## Residuals from last author



```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 877   70450198432 3.724 2009     1205      1     2.960
## 889   69949103890 4.473 2009     1205      1     3.733
## 992   79958058988 4.134 2011     1205      1     2.685
## 1070  84871070627 4.115 2012     1205      1     3.094
## 1089  84872909005 4.721 2012     1202      4     3.725
## 1100  84873811421 4.115 2012     1205      1     3.119
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6403 -0.7168 -0.0182  0.6212  3.7314
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1911      0.1416   8.41  <2e-16 ***
## LastAuthorFemale1  0.0207      0.0722   0.29   0.774
## Year1997          0.0371      0.1881   0.20   0.844
## Year1998          0.0117      0.1905   0.06   0.951
## Year1999         -0.2410      0.2006  -1.20   0.230
## Year2000         -0.1778      0.2002  -0.89   0.375
## Year2001         -0.1274      0.2002  -0.64   0.525
```

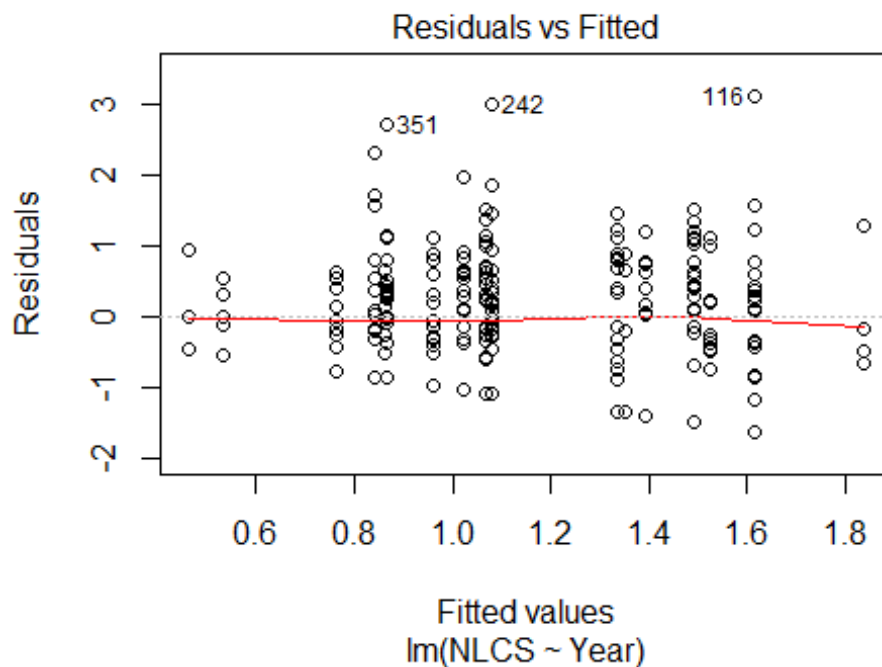


```

## Year2002          0.4284      0.2391      1.79      0.073 .
## Year2003          -0.1212      0.1964     -0.62      0.537
## Year2004          -0.1232      0.1775     -0.69      0.488
## Year2005           0.1776      0.1954      0.91      0.363
## Year2006          -0.1770      0.1744     -1.02      0.310
## Year2007          -0.2031      0.1742     -1.17      0.244
## Year2008          -0.3566      0.1837     -1.94      0.053 .
## Year2009          -0.4495      0.2185     -2.06      0.040 *
## Year2010          -0.5596      0.2344     -2.39      0.017 *
## Year2011           0.2348      0.2015      1.17      0.244
## Year2012          -0.1911      0.1922     -0.99      0.320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.924
## Multiple R-squared:  0.0569, Adjusted R-squared:  0.0358
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 73 weights are ~= 1. The remaining 706 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0663 0.8870 0.9410 0.9130 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.28e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 779"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1206"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   12    8    5   19   10   14   13   26   23   35   34   33   23   28
## 2011 2012

```

```
## 32 33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 8 4 4 12 6 12 10 21 20 30 31 30 16 19
## 2011 2012
## 27 31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 8 4 4 10 6 11 10 21 20 29 30 30 15 19
## 2011 2012
## 25 29
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```

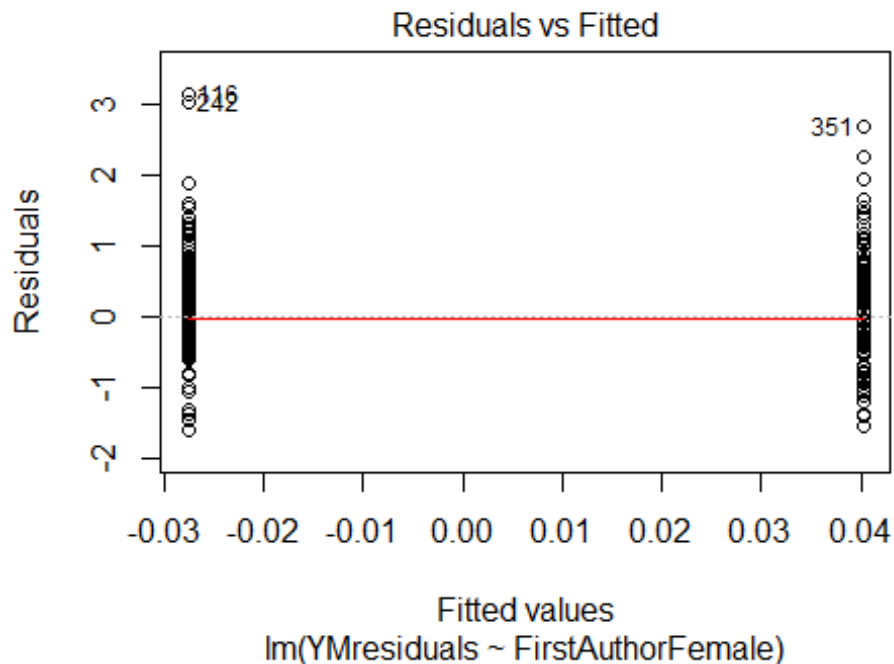


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6
## [1] "Female first author team size 2018 geometric mean: 1.60071019479117"
## [1] "Male first author team size 2018 geometric mean: 1.45889331870546"
```

```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 850, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.5687618032671"
## [1] "Male last author team size 2018 geometric mean: 1.48831945965861"

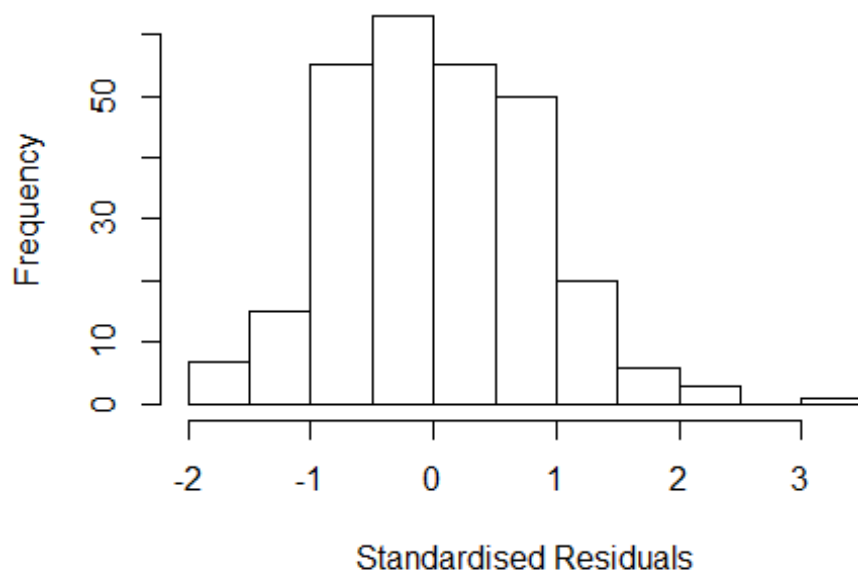
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 820, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
##      Year as factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.743  1          1.656
```

## LastAuthorFemale	2.475	1	1.573
## UniqueAuthors	4.780	4	1.216
## Year	7.292	16	1.064

## Residuals from first and last author and team size



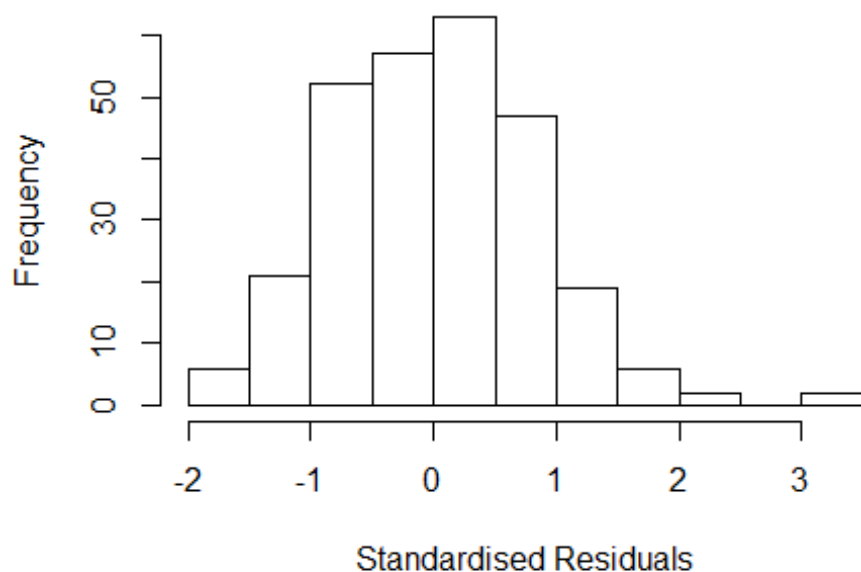
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 116 1842481625 4.725 2004      1206      1      3.065
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.72000 -0.61322 -0.00595  0.57036  3.06512
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4396    0.2959     1.49  0.13861
## FirstAuthorFemale1  0.0948    0.1529     0.62  0.53583
## LastAuthorFemale1 -0.0856    0.1555    -0.55  0.58254
## UniqueAuthors2    0.2699    0.1406     1.92  0.05606 .
## UniqueAuthors3    0.8008    0.2341     3.42  0.00073 ***
## UniqueAuthors4    0.5208    0.2472     2.11  0.03610 *
```

```

## UniqueAuthors5      0.2446      0.2713      0.90  0.36821
## Year1997             0.0947      0.3248      0.29  0.77076
## Year1998             0.8888      0.5600      1.59  0.11373
## Year1999             1.2173      0.5167      2.36  0.01925 *
## Year2000             0.0909      0.3503      0.26  0.79548
## Year2001             0.1995      0.3245      0.61  0.53936
## Year2002             0.5686      0.4429      1.28  0.20037
## Year2003             0.8555      0.3799      2.25  0.02517 *
## Year2004             0.9504      0.3699      2.57  0.01077 *
## Year2005             0.7722      0.3872      1.99  0.04718 *
## Year2006             1.0796      0.3573      3.02  0.00277 **
## Year2007             0.5038      0.3429      1.47  0.14303
## Year2008             0.4703      0.3399      1.38  0.16770
## Year2009             0.4176      0.3603      1.16  0.24755
## Year2010             0.1973      0.3700      0.53  0.59437
## Year2011             0.2423      0.3342      0.72  0.46920
## Year2012             0.3442      0.3361      1.02  0.30682
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.847
## Multiple R-squared:  0.186, Adjusted R-squared:  0.115
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.162  0.889   0.948   0.918   0.988   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.64e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.462 1          1.569
## LastAuthorFemale  2.178 1          1.476
## Year              1.782 16          1.018

```

## Residuals from first and last author



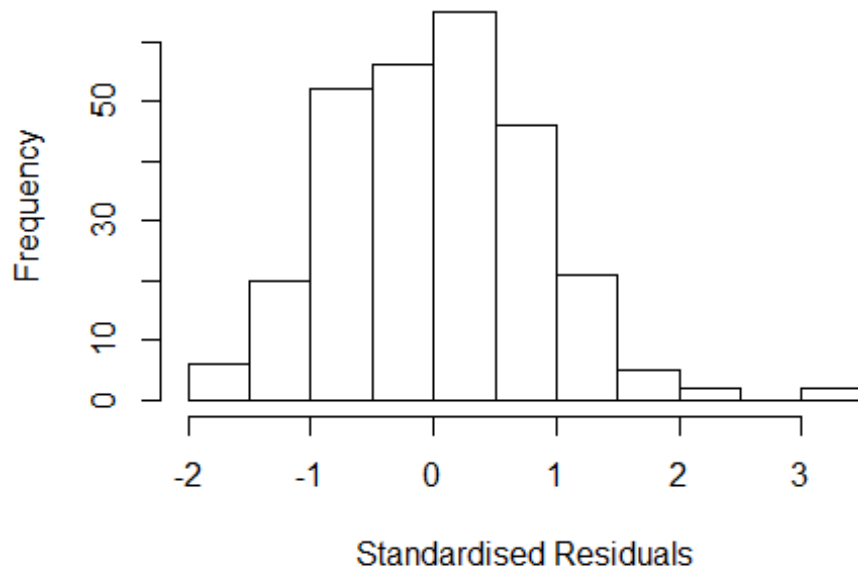
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 116  1842481625 4.725 2004    1206     1    3.256
## 242  84973622246 4.091 2008    1206     3    3.166
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6339 -0.6271  0.0109  0.5720  3.2557
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4392    0.2961   1.48  0.1393
## FirstAuthorFemale1 0.1872    0.1516   1.24  0.2178
## LastAuthorFemale1 -0.1434    0.1507  -0.95  0.3420
## Year1997         0.0908    0.3244   0.28  0.7798
## Year1998         0.9692    0.5965   1.62  0.1054
## Year1999         1.2611    0.5117   2.46  0.0144 *
## Year2000         0.4342    0.3419   1.27  0.2052
## Year2001         0.4074    0.3563   1.14  0.2539
## Year2002         0.9431    0.4143   2.28  0.0236 *
## Year2003         1.0490    0.3551   2.95  0.0034 **
## Year2004         1.0301    0.3761   2.74  0.0066 **
```

```

## Year2005          0.8558      0.3931      2.18      0.0304 *
## Year2006          1.1509      0.3619      3.18      0.0017 **
## Year2007          0.5740      0.3499      1.64      0.1021
## Year2008          0.4863      0.3333      1.46      0.1458
## Year2009          0.4302      0.3587      1.20      0.2315
## Year2010          0.2404      0.3690      0.65      0.5154
## Year2011          0.3188      0.3328      0.96      0.3389
## Year2012          0.4956      0.3346      1.48      0.1398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.839
## Multiple R-squared:  0.14,    Adjusted R-squared:  0.0791
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0985 0.8860 0.9480 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.64e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.37 1          1.171
## Year              1.37 16          1.010

```

## Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 116  1842481625 4.725 2004     1206      1      3.256
## 242  84973622246 4.091 2008     1206      3      3.166
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6477 -0.6591  0.0119  0.5622  3.2912
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4391     0.2962   1.48   0.1394
## FirstAuthorFemale1 0.0790     0.1150   0.69   0.4926
## Year1997          0.0863     0.3239   0.27   0.7900
## Year1998          0.9663     0.6048   1.60   0.1113
## Year1999          1.2853     0.5122   2.51   0.0127 *
## Year2000          0.3891     0.3334   1.17   0.2443
## Year2001          0.3959     0.3562   1.11   0.2675
## Year2002          0.9370     0.4165   2.25   0.0253 *
## Year2003          1.0395     0.3553   2.93   0.0037 **
## Year2004          0.9947     0.3745   2.66   0.0084 **
## Year2005          0.8761     0.3944   2.22   0.0272 *
```

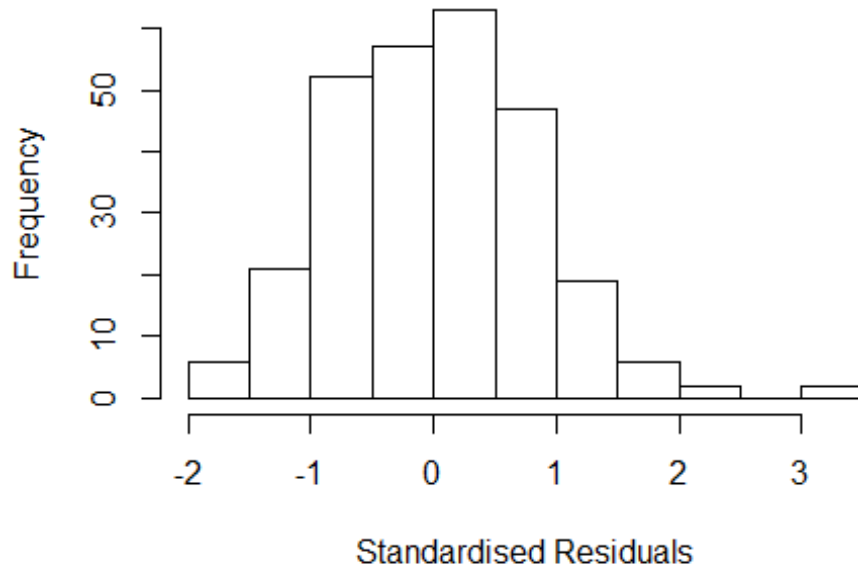


```

## Year2006          1.1296      0.3598      3.14      0.0019 **
## Year2007          0.5687      0.3505      1.62      0.1059
## Year2008          0.4803      0.3334      1.44      0.1509
## Year2009          0.4329      0.3592      1.21      0.2292
## Year2010          0.2200      0.3687      0.60      0.5513
## Year2011          0.3174      0.3330      0.95      0.3414
## Year2012          0.4734      0.3331      1.42      0.1565
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.838
## Multiple R-squared:  0.137, Adjusted R-squared:  0.0803
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 256 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0883 0.8860 0.9450 0.9110 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.64e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.222 1      1.105
## Year              1.222 16      1.006

```

## Residuals from last author



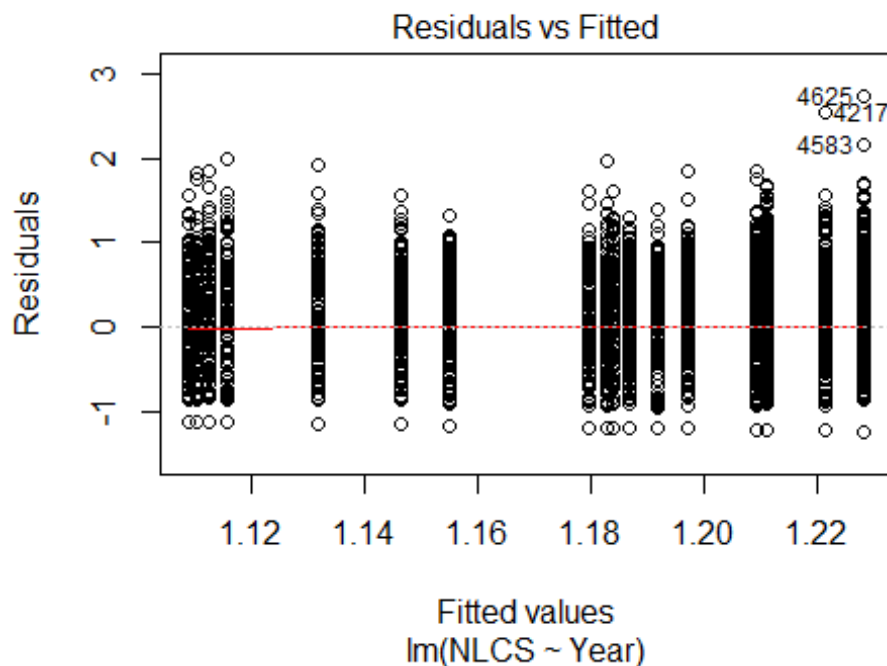
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 116  1842481625 4.725 2004      1206      1      3.256
## 242  84973622246 4.091 2008      1206      3      3.166
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.60950 -0.67734  0.00442  0.59228  3.25431
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.439043   0.296236   1.48   0.1395
## LastAuthorFemale1 0.000652   0.114594   0.01   0.9955
## Year1997        0.096431   0.325254   0.30   0.7671
## Year1998        0.974267   0.588195   1.66   0.0989 .
## Year1999        1.331601   0.495806   2.69   0.0077 **
## Year2000        0.395122   0.333198   1.19   0.2368
## Year2001        0.421731   0.355975   1.18   0.2372
## Year2002        0.945041   0.419699   2.25   0.0252 *
## Year2003        1.060351   0.354543   2.99   0.0031 **
## Year2004        1.031645   0.377116   2.74   0.0067 **
## Year2005        0.912999   0.390305   2.34   0.0201 *
```

```

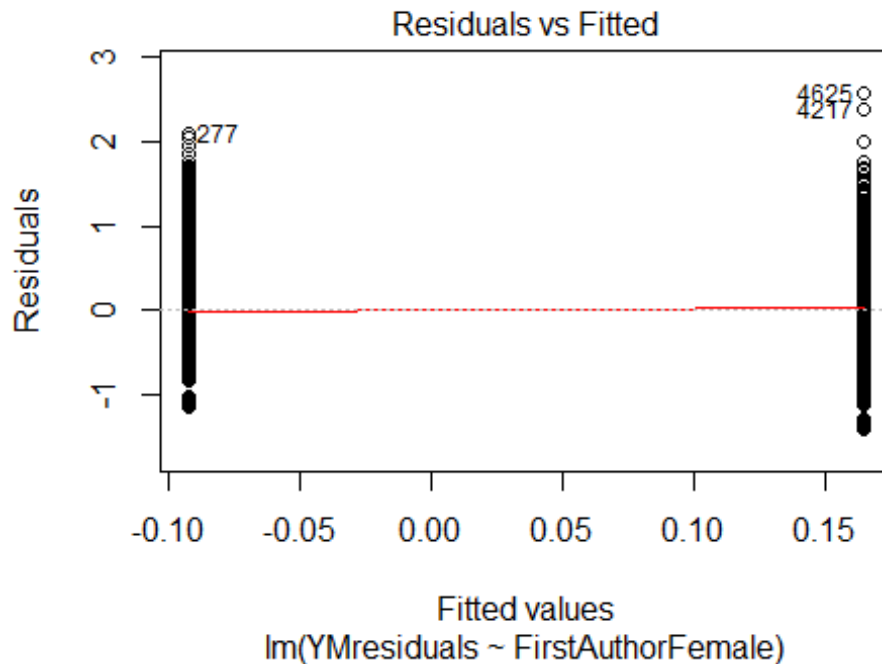
## Year2006      1.169803    0.364015    3.21    0.0015 **
## Year2007      0.592839    0.350623    1.69    0.0921 .
## Year2008      0.518972    0.331767    1.56    0.1190
## Year2009      0.476533    0.350226    1.36    0.1748
## Year2010      0.261714    0.369388    0.71    0.4793
## Year2011      0.349865    0.332841    1.05    0.2942
## Year2012      0.508278    0.334635    1.52    0.1300
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.836
## Multiple R-squared:  0.136, Adjusted R-squared:  0.0786
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0962 0.8840 0.9440 0.9110 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.64e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##      trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 275"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1207"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 173 199 166 178 206 201 199 162 174 191 310 315 321 302 381
## 2011 2012
## 332 348
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 136 139 133 138 161 158 161 124 144 150 260 274 287 263 324

```

```
## 2011 2012
## 294 300
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 132 136 131 132 158 152 158 122 141 148 256 266 275 261 318
## 2011 2012
## 286 297
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 2e-04
```

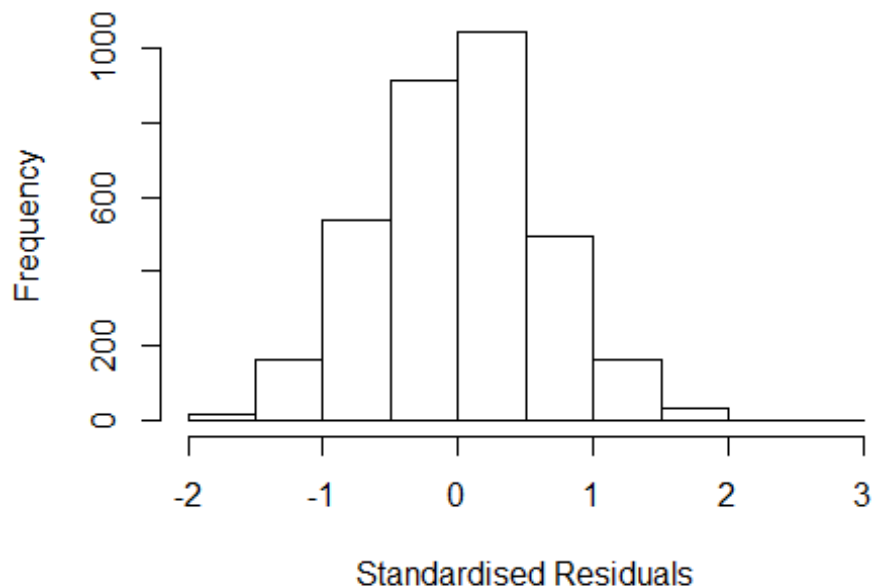


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.63, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 1.44800089634962"
## [1] "Male first author team size 2018 geometric mean: 1.19360172254183"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 9e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.42081426253395"
## [1] "Male last author team size 2018 geometric mean: 1.20956407219388"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.788 1          1.337
## LastAuthorFemale  1.665 1          1.290
## UniqueAuthors    1.242 4          1.027
## Year              1.145 16         1.004
```

## Residuals from first and last author and team size



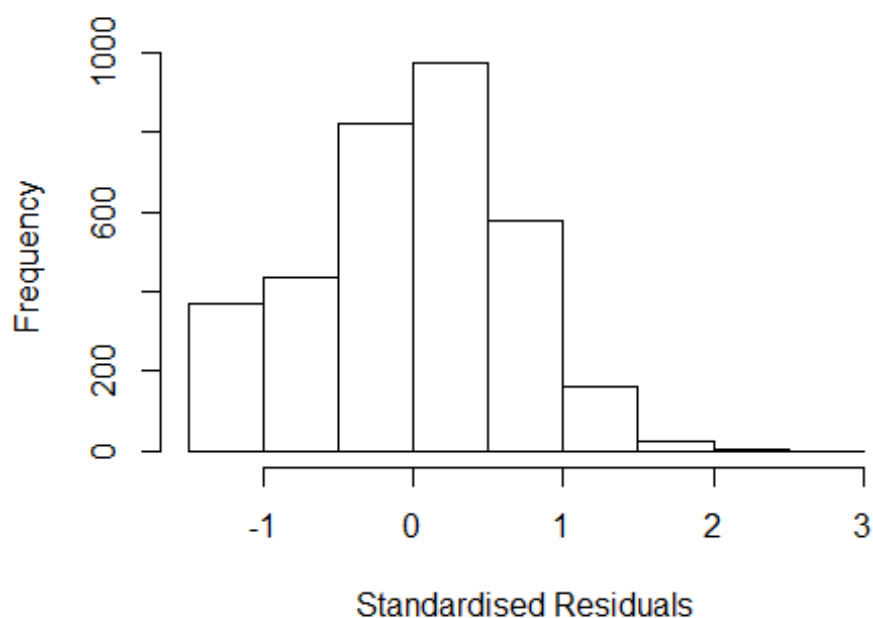
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4217 78751539841 3.763 2011      1202      3      2.547
## 4625 84855217343 3.954 2012      1207      2      2.743
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7254 -0.4195  0.0229  0.4169  2.7426
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.917060   0.065267   14.05 < 2e-16 ***
## FirstAuthorFemale1 0.155526   0.030059    5.17 2.4e-07 ***
## LastAuthorFemale1  0.049246   0.029801    1.65  0.099 .
## UniqueAuthors2    0.385891   0.031602   12.21 < 2e-16 ***
## UniqueAuthors3    0.558811   0.037142   15.05 < 2e-16 ***
## UniqueAuthors4    0.469643   0.046817   10.03 < 2e-16 ***
## UniqueAuthors5    0.581610   0.040875   14.23 < 2e-16 ***
## Year1997          -0.022595   0.087209   -0.26  0.796
## Year1998           0.048049   0.083113    0.58  0.563
```

```

## Year1999      0.064029    0.088319    0.72    0.469
## Year2000      0.037716    0.084859    0.44    0.657
## Year2001      0.018542    0.086388    0.21    0.830
## Year2002      0.024995    0.083707    0.30    0.765
## Year2003      0.049382    0.081204    0.61    0.543
## Year2004     -0.008992    0.082598   -0.11    0.913
## Year2005      0.049048    0.081063    0.61    0.545
## Year2006      0.007444    0.076168    0.10    0.922
## Year2007      0.070481    0.075921    0.93    0.353
## Year2008      0.058335    0.073331    0.80    0.426
## Year2009     -0.000598    0.073444   -0.01    0.994
## Year2010      0.074029    0.074424    0.99    0.320
## Year2011      0.094051    0.075794    1.24    0.215
## Year2012      0.089555    0.080244    1.12    0.264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.137, Adjusted R-squared:  0.132
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 287 weights are ~= 1. The remaining 3082 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0138 0.8540 0.9520 0.9050 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.011 1      1.418
## LastAuthorFemale 1.988 1      1.410
## Year      1.042 16      1.001

```

## Residuals from first and last author



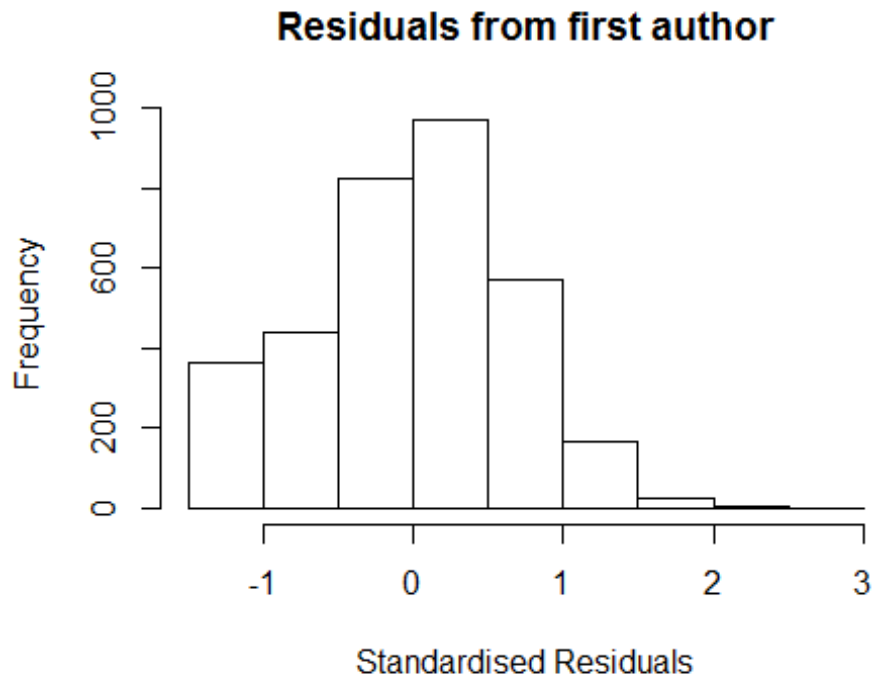
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4625 84855217343 3.954 2012    1207      2      2.569
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4065 -0.4690  0.0345  0.4588  2.5693
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02916    0.06991   14.72 < 2e-16 ***
## FirstAuthorFemale1 0.26354    0.03469    7.60 3.9e-14 ***
## LastAuthorFemale1 0.02521    0.03553    0.71  0.48
## Year1997        -0.05007    0.09508   -0.53  0.60
## Year1998         0.05632    0.09098    0.62  0.54
## Year1999         0.04801    0.09609    0.50  0.62
## Year2000        -0.02551    0.08842   -0.29  0.77
## Year2001         0.00169    0.09241    0.02  0.99
## Year2002         0.03453    0.08741    0.40  0.69
## Year2003         0.08854    0.08827    1.00  0.32
## Year2004        -0.01938    0.08997   -0.22  0.83
## Year2005         0.05247    0.08765    0.60  0.55
```



```

## Year2006          0.05343      0.08091      0.66      0.51
## Year2007          0.07630      0.08130      0.94      0.35
## Year2008          0.04425      0.07887      0.56      0.57
## Year2009          0.01538      0.07950      0.19      0.85
## Year2010          0.07315      0.08006      0.91      0.36
## Year2011          0.07386      0.08122      0.91      0.36
## Year2012          0.06682      0.08555      0.78      0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.676
## Multiple R-squared:  0.0422, Adjusted R-squared:  0.0371
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 252 weights are ~= 1. The remaining 3117 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.117  0.870   0.950   0.911   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1      1.014
## Year              1.029 16      1.001

```

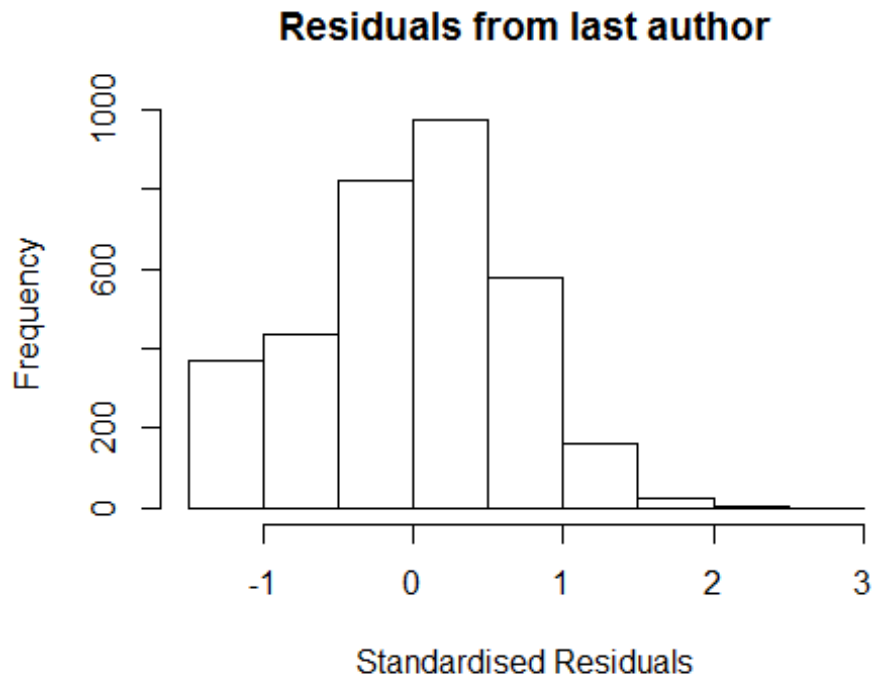


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4625 84855217343 3.954 2012    1207      2    2.569
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4008 -0.4712  0.0328  0.4561  2.5743
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03026    0.06996   14.73  <2e-16 ***
## FirstAuthorFemale1 0.28135    0.02480   11.34  <2e-16 ***
## Year1997      -0.04870    0.09505   -0.51    0.61
## Year1998       0.05644    0.09099    0.62    0.54
## Year1999       0.04848    0.09617    0.50    0.61
## Year2000      -0.02472    0.08843   -0.28    0.78
## Year2001       0.00147    0.09240    0.02    0.99
## Year2002       0.03626    0.08737    0.42    0.68
## Year2003       0.08923    0.08839    1.01    0.31
## Year2004      -0.01802    0.08996   -0.20    0.84
## Year2005       0.05264    0.08767    0.60    0.55
## Year2006       0.05387    0.08096    0.67    0.51
```

```

## Year2007          0.07694    0.08131    0.95    0.34
## Year2008          0.04617    0.07887    0.59    0.56
## Year2009          0.01546    0.07953    0.19    0.85
## Year2010          0.07404    0.08009    0.92    0.36
## Year2011          0.07559    0.08118    0.93    0.35
## Year2012          0.06814    0.08550    0.80    0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.676
## Multiple R-squared:  0.042, Adjusted R-squared:  0.0372
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 258 weights are ~= 1. The remaining 3111 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.115  0.870  0.950  0.911  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1          1.008
## Year              1.016 16          1.000

```



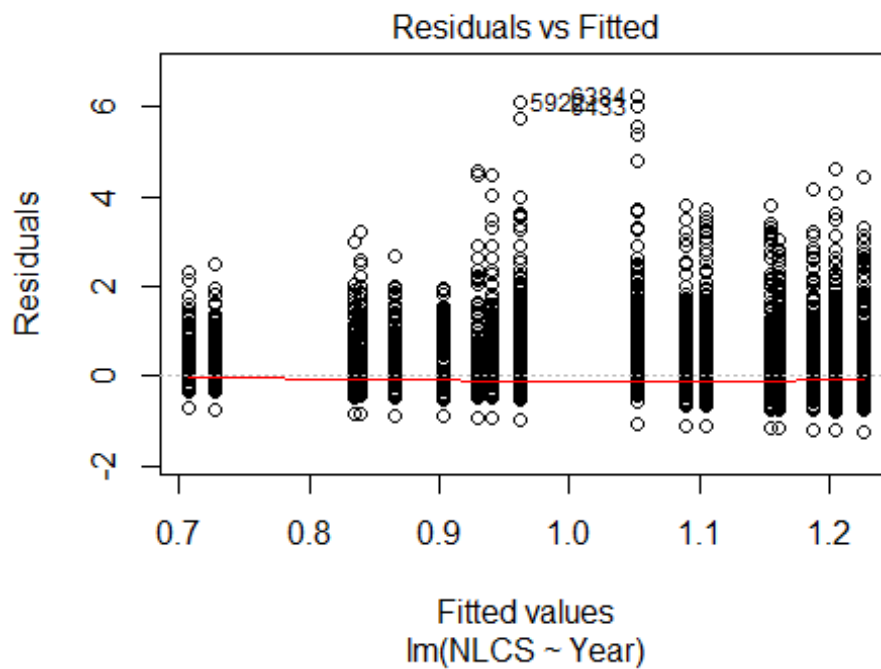
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4625 84855217343 3.954 2012    1207    2    2.569
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3645 -0.4734  0.0325  0.4709  2.6093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05313    0.07040   14.96  <2e-16 ***
## LastAuthorFemale1 0.21785    0.02517    8.66  <2e-16 ***
## Year1997       -0.05327    0.09630   -0.55    0.58
## Year1998        0.06432    0.09201    0.70    0.48
## Year1999        0.05104    0.09630    0.53    0.60
## Year2000       -0.01910    0.08891   -0.21    0.83
## Year2001        0.00715    0.09347    0.08    0.94
## Year2002        0.03129    0.08921    0.35    0.73
## Year2003        0.09353    0.08846    1.06    0.29
## Year2004       -0.02492    0.09063   -0.27    0.78
## Year2005        0.05909    0.08899    0.66    0.51
## Year2006        0.07510    0.08158    0.92    0.36
```

```

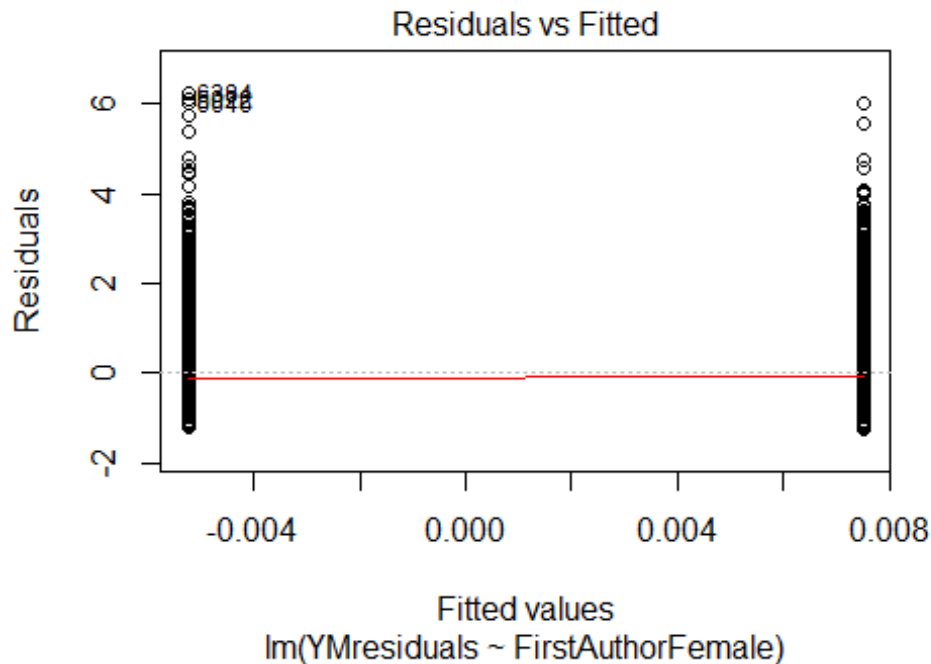
## Year2007      0.08540      0.08208      1.04      0.30
## Year2008      0.04612      0.07933      0.58      0.56
## Year2009      0.02988      0.08015      0.37      0.71
## Year2010      0.07915      0.08088      0.98      0.33
## Year2011      0.07566      0.08184      0.92      0.36
## Year2012      0.07370      0.08630      0.85      0.39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.688
## Multiple R-squared:  0.0259, Adjusted R-squared:  0.021
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 247 weights are ~= 1. The remaining 3122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.118  0.872  0.951   0.913  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3369"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1208"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 234 257 275 268 252 238 448 382 438 483 466 581 487 369 320
## 2011 2012
## 576 734
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 202 228 242 210 192 381 337 386 423 408 517 418 320 271
## 2011 2012

```

```
## 499 642
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 202 226 240 209 191 381 337 382 420 407 515 416 320 271
## 2011 2012
## 497 638
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 300, df = 16, p-value <2e-16
```

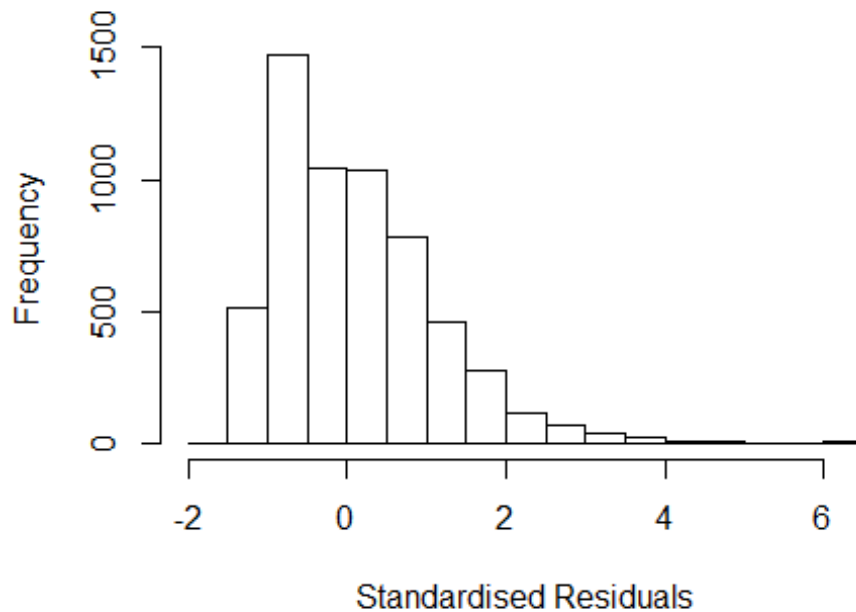


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.2, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 1.06447400079714"
## [1] "Male first author team size 2018 geometric mean: 1.03063765010677"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 55000, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.06120608924178"
## [1] "Male last author team size 2018 geometric mean: 1.033659622802"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 55000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 14.558 1          3.816
## LastAuthorFemale  14.517 1          3.810
## UniqueAuthors     1.106 4          1.013
## Year              1.063 16          1.002
```

## Residuals from first and last author and team size



```
## [1] "List of 153 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3    33746446640 3.452 1996    1208      1    2.702
## 6    60950433344 4.062 1996    1208      1    3.312
## 133  85015910893 3.345 1996    1208      1    2.595
## 240  60949595059 3.564 1997    1208      1    2.771
## 552  61249601308 3.222 1998    1202      4    2.543
## 1080 61249468539 3.822 2000    1208      3    3.049
## 1545 35448994621 3.955 2002    1208      1    2.906
## 1548 60949487473 3.855 2002    1208      1    2.806
## 1588 60950373677 3.920 2002    1208      2    2.871
## 1746 61249391353 4.222 2002    1208      1    3.173
## 1751 61249678334 3.746 2002    1208      1    2.674
## 1785 48949111059 4.224 2002    1208      2    3.152
## 1793 61949171102 3.665 2002    1208      2    2.616
## 1851 60949941202 3.850 2002    1202      3    2.801
## 1994 46449113576 4.314 2003    1208      1    3.256
## 1999 60949679561 3.954 2003    1208      1    2.872
## 2024 63449141239 3.780 2003    1208      1    2.722
## 2027 27844509137 4.429 2003    1208      2    3.371
## 2028 33645366634 4.044 2003    1208      2    2.962
## 2181 60950014779 3.580 2003    1208      1    2.522
## 2203 64249135095 4.034 2003    1208      1    2.976
## 2223 34248348215 5.357 2003    1208      2    4.299
## 2378 50549091357 4.213 2004    1208      1    3.128
## 2389 60950006630 3.947 2004    1208      1    2.862
## 2392 60950365720 3.622 2004    1208      1    2.537
```



##	2395	60950504818	4.213	2004	1208	1	3.128
##	2398	60950612949	3.947	2004	1208	1	2.862
##	2400	61049193685	3.622	2004	1208	1	2.513
##	2422	19644393451	3.827	2004	1208	2	2.742
##	2444	60950717875	3.827	2004	1208	2	2.742
##	2587	49849105643	4.367	2004	1208	1	3.258
##	2589	60949290754	4.041	2004	1208	1	2.956
##	2590	60949341250	3.847	2004	1208	1	2.762
##	2618	67651236094	3.739	2004	1208	1	2.630
##	2624	70450031636	3.739	2004	1208	1	2.630
##	2640	41749085870	5.674	2004	1208	2	4.589
##	2641	52949127270	3.675	2004	1208	2	2.590
##	2645	61149315889	4.521	2004	1208	2	3.412
##	2669	67649687426	3.675	2004	1208	2	2.590
##	2828	34347305000	5.282	2005	1208	1	4.190
##	2865	61249405456	5.282	2005	1208	1	4.190
##	2886	67649786656	4.683	2005	1208	1	3.615
##	3134	70350414060	4.001	2005	1208	2	2.933
##	3135	79957120427	3.708	2005	1208	2	2.640
##	3145	36849063234	5.820	2005	1208	1	4.752
##	3149	60950558911	4.683	2005	1208	1	3.591
##	3166	64949203782	3.688	2005	1208	1	2.620
##	3184	85007789945	4.358	2005	1208	1	3.290
##	3185	85012571734	4.822	2005	1208	1	3.730
##	3186	85014004514	3.822	2005	1208	1	2.754
##	3323	43249162424	3.908	2006	1208	1	2.859
##	3329	60949501532	3.786	2006	1208	1	2.761
##	3337	61149220396	4.126	2006	1208	1	3.101
##	3556	70349380046	4.224	2006	1208	1	3.199
##	3623	34247870091	4.486	2006	1208	1	3.017
##	3637	60950624832	4.486	2006	1208	1	3.461
##	3641	61149487712	4.564	2006	1208	1	3.515
##	3645	67649756861	4.317	2006	1208	1	3.292
##	3646	67649803123	3.908	2006	1208	1	2.859
##	3665	60949428823	4.945	2006	1208	2	3.896
##	3683	70449761232	4.357	2006	1208	2	3.308
##	3784	43249143287	4.082	2007	1208	1	3.093
##	3799	57749162256	4.449	2007	1208	1	3.484
##	3802	57749166879	3.939	2007	1208	1	2.950
##	3818	61149122409	4.555	2007	1208	1	3.590
##	3837	67650088013	4.837	2007	1208	1	3.872
##	3857	60949396166	3.834	2007	1208	2	2.869
##	3888	61949171840	3.629	2007	1208	2	2.664
##	4098	61049174750	4.837	2007	1208	1	3.848
##	4099	61049252344	3.939	2007	1208	1	2.974
##	4100	61049380407	4.335	2007	1208	1	2.635
##	4165	48749117113	4.214	2007	1208	1	3.249
##	4177	61449324554	3.939	2007	1208	1	2.974
##	4186	70350522140	4.654	2007	1208	1	3.689
##	4220	61149608982	3.629	2007	1208	2	2.640

##	4370	60950688715	3.598	2008	1208	1	2.616
##	4406	57749135069	4.149	2008	1208	1	3.191
##	4432	61149730225	4.566	2008	1208	1	3.608
##	4642	45349083255	3.630	2008	1208	2	2.648
##	4697	61449322260	4.300	2008	1208	1	3.318
##	4698	61449425485	4.902	2008	1208	1	3.944
##	4713	79960301345	4.149	2008	1208	1	3.167
##	4722	84977516653	3.598	2008	1208	1	2.640
##	4728	85014203193	3.984	2008	1208	1	3.026
##	4862	75849129999	3.591	2009	1208	1	2.817
##	4981	70350018180	3.591	2009	1208	1	2.793
##	5018	68949127004	4.249	2009	1208	1	3.451
##	5097	65849109286	3.306	2009	1208	1	2.532
##	5099	65849275932	3.837	2009	1208	1	3.039
##	5103	77950760036	3.306	2009	1208	1	2.532
##	5106	84865299965	3.591	2009	1208	1	2.793
##	5107	84873377619	4.425	2009	1208	1	3.651
##	5110	84909053994	3.837	2009	1208	1	3.039
##	5111	84918906073	3.306	2009	1208	1	2.532
##	5117	85008922077	5.433	2009	1208	1	4.659
##	5122	85009553339	3.306	2009	1208	1	2.508
##	5131	85009635345	3.306	2009	1208	1	2.508
##	5135	85014598252	3.306	2009	1208	1	2.532
##	5147	77949694173	4.988	2009	1208	2	4.190
##	5371	79957915084	5.520	2010	1208	1	4.714
##	5372	79957915644	3.832	2010	1208	1	3.050
##	5479	77950130289	5.426	2010	1208	1	4.644
##	5480	77950168892	3.586	2010	1208	1	2.780
##	5517	79955439227	3.302	2010	1208	1	2.520
##	5589	84857698666	4.563	2011	1208	2	3.780
##	5598	84855915732	4.594	2011	1208	4	3.835
##	5599	84855916060	4.298	2011	1208	4	3.515
##	5600	84855927603	3.619	2011	1208	4	2.860
##	5728	80051973513	3.887	2011	1208	1	3.104
##	5729	80052003294	3.491	2011	1208	1	2.732
##	5742	84861078242	4.221	2011	1208	1	3.438
##	5816	79958246980	3.887	2011	1208	1	3.104
##	5817	79958252976	3.887	2011	1208	1	3.128
##	5824	80052376495	3.491	2011	1208	1	2.708
##	5866	79960771503	3.887	2011	1208	1	3.128
##	5913	79958063974	3.323	2011	1203	6	2.564
##	5922	79957873219	7.067	2011	1208	1	6.308
##	5994	79952748470	4.221	2011	1208	1	3.462
##	5998	79955375481	4.511	2011	1208	1	3.728
##	6000	79955410959	6.705	2011	1208	1	5.946
##	6007	79956198385	4.972	2011	1208	2	4.189
##	6057	79960875954	3.491	2011	1208	1	2.708
##	6076	84862022685	3.611	2011	1205	2	2.828
##	6166	84870369563	4.382	2012	1208	1	3.549
##	6169	84868627182	3.937	2012	1208	1	3.128

```

## 6174 84871098188 3.391 2012      1208      1      2.558
## 6189 84876955971 4.759 2012      1208      1      3.950
## 6190 84876957904 4.382 2012      1208      1      3.573
## 6191 84876972474 5.865 2012      1208      1      5.056
## 6193 84877007167 5.865 2012      1208      1      5.056
## 6222 84871243321 4.302 2012      1208      2      3.493
## 6330 84868306956 3.686 2012      1208      2      2.877
## 6378 84867794028 4.759 2012      1208      1      3.926
## 6384 84866140479 7.287 2012      1208      2      6.478
## 6386 84868655179 3.559 2012      1208      2      2.750
## 6387 84868667202 3.962 2012      1208      2      3.129
## 6423 84866632349 3.391 2012      1208      1      2.582
## 6425 84866948227 6.455 2012      1208      1      5.646
## 6433 84874479222 7.070 2012      1208      1      6.237
## 6542 84861079400 3.962 2012      1208      2      3.129
## 6565 84864673383 3.391 2012      1208      1      2.558
## 6566 84866065375 3.391 2012      1208      1      2.558
## 6587 84859957258 3.520 2012      1202      4      2.711
## 6616 84859326587 7.070 2012      1208      1      6.261
## 6636 84859765129 6.624 2012      1208      1      5.791
## 6655 84872909005 4.721 2012      1202      4      3.552
## 6676 84858641316 3.391 2012      1208      1      2.558
## 6679 84859471937 3.391 2012      1208      1      2.582
## 6681 84860350141 4.382 2012      1208      1      3.573
## 6684 84861541436 3.937 2012      1208      1      3.104
## 6689 84856425818 5.835 2012      1208      2      5.002
## 6691 84856507096 3.559 2012      1208      2      2.750
## 6790 84865745109 3.962 2012      1208      2      3.129
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7194 -0.7744 -0.0332  0.6766  6.4777
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.75021    0.05709   13.14 < 2e-16 ***
## FirstAuthorFemale1  0.08380    0.09589    0.87  0.38223
## LastAuthorFemale1 -0.05985    0.09574   -0.63  0.53192
## UniqueAuthors2     0.36017    0.07629    4.72  2.4e-06 ***
## UniqueAuthors3     0.65122    0.20274    3.21  0.00133 **
## UniqueAuthors4     0.81457    0.46688    1.74  0.08109 .
## UniqueAuthors5    -0.03616    0.29938   -0.12  0.90388
## Year1997         0.04235    0.07860    0.54  0.59008

```

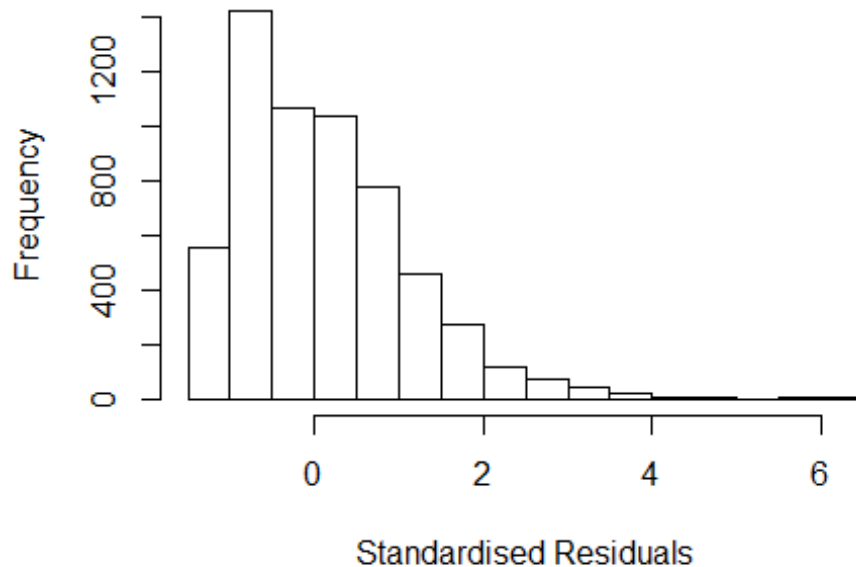
```

## Year1998      -0.07096      0.07188      -0.99      0.32358
## Year1999      -0.08629      0.07147      -1.21      0.22729
## Year2000       0.02282      0.07413       0.31      0.75825
## Year2001       0.10923      0.07841       1.39      0.16367
## Year2002       0.29832      0.07738       3.86      0.00012 ***
## Year2003       0.30784      0.07982       3.86      0.00012 ***
## Year2004       0.33449      0.07603       4.40      1.1e-05 ***
## Year2005       0.31793      0.07535       4.22      2.5e-05 ***
## Year2006       0.27498      0.07318       3.76      0.00017 ***
## Year2007       0.21459      0.07220       2.97      0.00297 **
## Year2008       0.20794      0.07747       2.68      0.00729 **
## Year2009       0.02417      0.07768       0.31      0.75566
## Year2010       0.03183      0.07951       0.40      0.68898
## Year2011       0.00903      0.07266       0.12      0.90113
## Year2012       0.05907      0.06982       0.85      0.39753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.901
## Multiple R-squared:  0.0291, Adjusted R-squared:  0.0254
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 16 observations
## c(1850,2216,2654,4345,4560,4645,5023,5089,5252,5254,5420,5453,5460,5613,5628,
## 5675)
## are outliers with |weight| = 0 ( < 1.7e-05);
## 523 weights are ~ 1. The remaining 5313 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0002 0.8800 0.9320 0.8960 0.9800 0.9990
## Algorithmic parameters:
## tuning.chi      bb      tuning.psi      refine.tol
## 1.55e+00      5.00e-01      4.69e+00      1.00e-07
## rel.tol      solve.tol      eps.outlier      eps.x
## 1.00e-07      1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
## 500      50      2      1      1000      200
## trace.lev      mts      compute.rd
## 0      1000      0
## psi      subsampling      cov
## "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
## factors"
## GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 18.053 1 4.249

```

## LastAuthorFemale	18.040	1	4.247
## Year	1.017	16	1.001

### Residuals from first and last author



```
## [1] "List of 154 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3    33746446640 3.452 1996    1208      1    2.682
## 6    60950433344 4.062 1996    1208      1    3.292
## 133  85015910893 3.345 1996    1208      1    2.575
## 240  60949595059 3.564 1997    1208      1    2.747
## 552  61249601308 3.222 1998    1202      4    2.528
## 1080 61249468539 3.822 2000    1208      3    3.037
## 1545 35448994621 3.955 2002    1208      1    2.889
## 1548 60949487473 3.855 2002    1208      1    2.789
## 1588 60950373677 3.920 2002    1208      2    2.854
## 1746 61249391353 4.222 2002    1208      1    3.156
## 1751 61249678334 3.746 2002    1208      1    2.655
## 1785 48949111059 4.224 2002    1208      2    3.133
## 1793 61949171102 3.665 2002    1208      2    2.599
## 1851 60949941202 3.850 2002    1202      3    2.784
## 1994 46449113576 4.314 2003    1208      1    3.241
## 1999 60949679561 3.954 2003    1208      1    2.856
## 2024 63449141239 3.780 2003    1208      1    2.707
## 2027 27844509137 4.429 2003    1208      2    3.356
## 2028 33645366634 4.044 2003    1208      2    2.946
## 2181 60950014779 3.580 2003    1208      1    2.507
## 2203 64249135095 4.034 2003    1208      1    2.961
## 2220 33749432390 3.807 2003    1208      2    2.683
```

##	2223	34248348215	5.357	2003	1208	2	4.284
##	2378	50549091357	4.213	2004	1208	1	3.120
##	2389	60950006630	3.947	2004	1208	1	2.854
##	2392	60950365720	3.622	2004	1208	1	2.529
##	2395	60950504818	4.213	2004	1208	1	3.120
##	2398	60950612949	3.947	2004	1208	1	2.854
##	2400	61049193685	3.622	2004	1208	1	2.504
##	2422	19644393451	3.827	2004	1208	2	2.734
##	2444	60950717875	3.827	2004	1208	2	2.734
##	2587	49849105643	4.367	2004	1208	1	3.249
##	2589	60949290754	4.041	2004	1208	1	2.948
##	2590	60949341250	3.847	2004	1208	1	2.754
##	2618	67651236094	3.739	2004	1208	1	2.621
##	2624	70450031636	3.739	2004	1208	1	2.621
##	2640	41749085870	5.674	2004	1208	2	4.581
##	2641	52949127270	3.675	2004	1208	2	2.582
##	2645	61149315889	4.521	2004	1208	2	3.403
##	2669	67649687426	3.675	2004	1208	2	2.582
##	2828	34347305000	5.282	2005	1208	1	4.180
##	2865	61249405456	5.282	2005	1208	1	4.180
##	2886	67649786656	4.683	2005	1208	1	3.606
##	3134	70350414060	4.001	2005	1208	2	2.924
##	3135	79957120427	3.708	2005	1208	2	2.631
##	3145	36849063234	5.820	2005	1208	1	4.743
##	3149	60950558911	4.683	2005	1208	1	3.581
##	3166	64949203782	3.688	2005	1208	1	2.611
##	3184	85007789945	4.358	2005	1208	1	3.281
##	3185	85012571734	4.822	2005	1208	1	3.720
##	3186	85014004514	3.822	2005	1208	1	2.745
##	3323	43249162424	3.908	2006	1208	1	2.840
##	3329	60949501532	3.786	2006	1208	1	2.744
##	3337	61149220396	4.126	2006	1208	1	3.084
##	3556	70349380046	4.224	2006	1208	1	3.182
##	3623	34247870091	4.486	2006	1208	1	3.393
##	3637	60950624832	4.486	2006	1208	1	3.444
##	3641	61149487712	4.564	2006	1208	1	3.496
##	3645	67649756861	4.317	2006	1208	1	3.275
##	3646	67649803123	3.908	2006	1208	1	2.840
##	3665	60949428823	4.945	2006	1208	2	3.877
##	3683	70449761232	4.357	2006	1208	2	3.289
##	3784	43249143287	4.082	2007	1208	1	3.080
##	3799	57749162256	4.449	2007	1208	1	3.472
##	3802	57749166879	3.939	2007	1208	1	2.937
##	3818	61149122409	4.555	2007	1208	1	3.578
##	3837	67650088013	4.837	2007	1208	1	3.860
##	3857	60949396166	3.834	2007	1208	2	2.857
##	3888	61949171840	3.629	2007	1208	2	2.652
##	4098	61049174750	4.837	2007	1208	1	3.835
##	4099	61049252344	3.939	2007	1208	1	2.962
##	4100	61049380407	4.335	2007	1208	1	3.307

##	4165	48749117113	4.214	2007	1208	1	3.237
##	4177	61449324554	3.939	2007	1208	1	2.962
##	4186	70350522140	4.654	2007	1208	1	3.677
##	4220	61149608982	3.629	2007	1208	2	2.627
##	4370	60950688715	3.598	2008	1208	1	2.603
##	4406	57749135069	4.149	2008	1208	1	3.179
##	4432	61149730225	4.566	2008	1208	1	3.596
##	4642	45349083255	3.630	2008	1208	2	2.635
##	4697	61449322260	4.300	2008	1208	1	3.305
##	4698	61449425485	4.902	2008	1208	1	3.932
##	4713	79960301345	4.149	2008	1208	1	3.154
##	4722	84977516653	3.598	2008	1208	1	2.628
##	4728	85014203193	3.984	2008	1208	1	3.014
##	4862	75849129999	3.591	2009	1208	1	2.805
##	4981	70350018180	3.591	2009	1208	1	2.780
##	5018	68949127004	4.249	2009	1208	1	3.438
##	5097	65849109286	3.306	2009	1208	1	2.520
##	5099	65849275932	3.837	2009	1208	1	3.026
##	5103	77950760036	3.306	2009	1208	1	2.520
##	5106	84865299965	3.591	2009	1208	1	2.780
##	5107	84873377619	4.425	2009	1208	1	3.639
##	5110	84909053994	3.837	2009	1208	1	3.026
##	5111	84918906073	3.306	2009	1208	1	2.520
##	5117	85008922077	5.433	2009	1208	1	4.647
##	5135	85014598252	3.306	2009	1208	1	2.520
##	5147	77949694173	4.988	2009	1208	2	4.177
##	5371	79957915084	5.520	2010	1208	1	4.693
##	5372	79957915644	3.832	2010	1208	1	3.030
##	5479	77950130289	5.426	2010	1208	1	4.624
##	5480	77950168892	3.586	2010	1208	1	2.759
##	5517	79955439227	3.302	2010	1208	1	2.500
##	5589	84857698666	4.563	2011	1208	2	3.764
##	5598	84855915732	4.594	2011	1208	4	3.820
##	5599	84855916060	4.298	2011	1208	4	3.499
##	5600	84855927603	3.619	2011	1208	4	2.845
##	5728	80051973513	3.887	2011	1208	1	3.088
##	5729	80052003294	3.491	2011	1208	1	2.717
##	5742	84861078242	4.221	2011	1208	1	3.422
##	5816	79958246980	3.887	2011	1208	1	3.088
##	5817	79958252976	3.887	2011	1208	1	3.113
##	5824	80052376495	3.491	2011	1208	1	2.692
##	5866	79960771503	3.887	2011	1208	1	3.113
##	5913	79958063974	3.323	2011	1203	6	2.549
##	5922	79957873219	7.067	2011	1208	1	6.293
##	5994	79952748470	4.221	2011	1208	1	3.447
##	5998	79955375481	4.511	2011	1208	1	3.712
##	6000	79955410959	6.705	2011	1208	1	5.931
##	6007	79956198385	4.972	2011	1208	2	4.173
##	6057	79960875954	3.491	2011	1208	1	2.692
##	6076	84862022685	3.611	2011	1205	2	2.812

```

## 6166 84870369563 4.382 2012      1208      1      3.534
## 6169 84868627182 3.937 2012      1208      1      3.114
## 6174 84871098188 3.391 2012      1208      1      2.543
## 6189 84876955971 4.759 2012      1208      1      3.936
## 6190 84876957904 4.382 2012      1208      1      3.559
## 6191 84876972474 5.865 2012      1208      1      5.042
## 6193 84877007167 5.865 2012      1208      1      5.042
## 6222 84871243321 4.302 2012      1208      2      3.479
## 6330 84868306956 3.686 2012      1208      2      2.863
## 6378 84867794028 4.759 2012      1208      1      3.911
## 6384 84866140479 7.287 2012      1208      2      6.464
## 6386 84868655179 3.559 2012      1208      2      2.736
## 6387 84868667202 3.962 2012      1208      2      3.114
## 6423 84866632349 3.391 2012      1208      1      2.568
## 6425 84866948227 6.455 2012      1208      1      5.632
## 6433 84874479222 7.070 2012      1208      1      6.222
## 6542 84861079400 3.962 2012      1208      2      3.114
## 6544 84864030671 3.567 2012      1208      2      2.769
## 6565 84864673383 3.391 2012      1208      1      2.543
## 6566 84866065375 3.391 2012      1208      1      2.543
## 6587 84859957258 3.520 2012      1202      4      2.697
## 6616 84859326587 7.070 2012      1208      1      6.247
## 6636 84859765129 6.624 2012      1208      1      5.776
## 6655 84872909005 4.721 2012      1202      4      3.898
## 6676 84858641316 3.391 2012      1208      1      2.543
## 6679 84859471937 3.391 2012      1208      1      2.568
## 6681 84860350141 4.382 2012      1208      1      3.559
## 6684 84861541436 3.937 2012      1208      1      3.089
## 6685 84864326825 3.391 2012      1208      1      2.568
## 6689 84856425818 5.835 2012      1208      2      4.987
## 6691 84856507096 3.559 2012      1208      2      2.736
## 6790 84865745109 3.962 2012      1208      2      3.114
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1436 -0.7864 -0.0363  0.6855  6.4637
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76956    0.05646   13.63 < 2e-16 ***
## FirstAuthorFemale1 0.05069    0.10729    0.47  0.63663
## LastAuthorFemale1 -0.02567    0.10725   -0.24  0.81086
## Year1997         0.04721    0.07831    0.60  0.54658
## Year1998        -0.07508    0.07137   -1.05  0.29282
## Year1999        -0.09602    0.07151   -1.34  0.17945

```

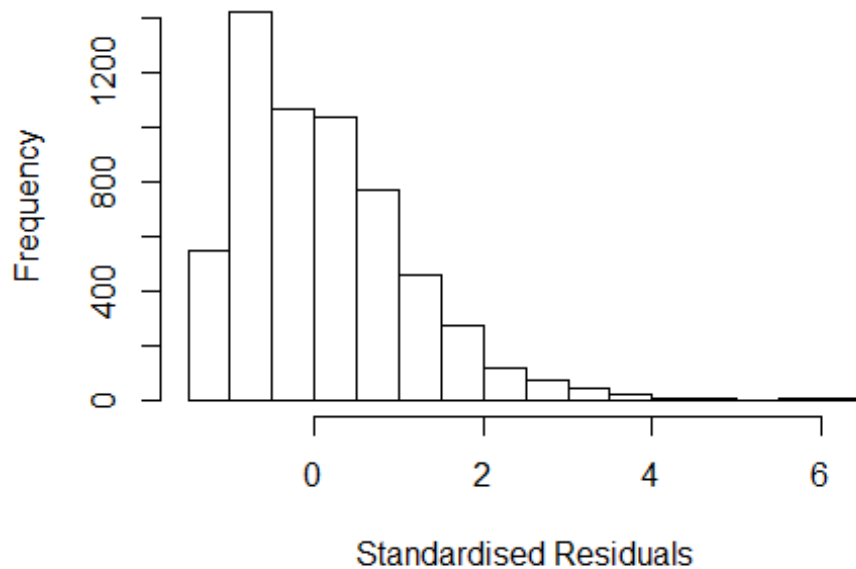


```

## Year2000      0.01589      0.07340      0.22  0.82862
## Year2001      0.10395      0.07851      1.32  0.18554
## Year2002      0.29595      0.07750      3.82  0.00014 ***
## Year2003      0.30366      0.07965      3.81  0.00014 ***
## Year2004      0.32333      0.07553      4.28  1.9e-05 ***
## Year2005      0.30702      0.07510      4.09  4.4e-05 ***
## Year2006      0.27293      0.07245      3.77  0.00017 ***
## Year2007      0.20743      0.07202      2.88  0.00399 **
## Year2008      0.20052      0.07656      2.62  0.00884 **
## Year2009      0.01686      0.07728      0.22  0.82726
## Year2010      0.03207      0.07924      0.40  0.68567
## Year2011      0.00408      0.07264      0.06  0.95521
## Year2012      0.05370      0.06990      0.77  0.44241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.905
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.0189
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 16 observations
## c(1850,2216,2654,4345,4560,4645,5023,5089,5252,5254,5420,5453,5460,5613,5628,
## 5675)
## are outliers with |weight| = 0 ( < 1.7e-05);
## 507 weights are ~= 1. The remaining 5329 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 0.8780 0.9300 0.8960 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.006 1      1.003
## Year      1.006 16      1.000

```

## Residuals from first author



```
## [1] "List of 154 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3      33746446640 3.452 1996      1208      1      2.682
## 6      60950433344 4.062 1996      1208      1      3.292
## 133    85015910893 3.345 1996      1208      1      2.575
## 240    60949595059 3.564 1997      1208      1      2.747
## 552    61249601308 3.222 1998      1202      4      2.528
## 1080   61249468539 3.822 2000      1208      3      3.037
## 1545   35448994621 3.955 2002      1208      1      2.889
## 1548   60949487473 3.855 2002      1208      1      2.789
## 1588   60950373677 3.920 2002      1208      2      2.854
## 1746   61249391353 4.222 2002      1208      1      3.156
## 1751   61249678334 3.746 2002      1208      1      2.655
## 1785   48949111059 4.224 2002      1208      2      3.133
## 1793   61949171102 3.665 2002      1208      2      2.599
## 1851   60949941202 3.850 2002      1202      3      2.784
## 1994   46449113576 4.314 2003      1208      1      3.241
## 1999   60949679561 3.954 2003      1208      1      2.856
## 2024   63449141239 3.780 2003      1208      1      2.707
## 2027   27844509137 4.429 2003      1208      2      3.356
## 2028   33645366634 4.044 2003      1208      2      2.946
## 2181   60950014779 3.580 2003      1208      1      2.507
## 2203   64249135095 4.034 2003      1208      1      2.961
## 2220   33749432390 3.807 2003      1208      2      2.683
## 2223   34248348215 5.357 2003      1208      2      4.284
## 2378   50549091357 4.213 2004      1208      1      3.120
## 2389   60950006630 3.947 2004      1208      1      2.854
```

##	2392	60950365720	3.622	2004	1208	1	2.529
##	2395	60950504818	4.213	2004	1208	1	3.120
##	2398	60950612949	3.947	2004	1208	1	2.854
##	2400	61049193685	3.622	2004	1208	1	2.504
##	2422	19644393451	3.827	2004	1208	2	2.734
##	2444	60950717875	3.827	2004	1208	2	2.734
##	2587	49849105643	4.367	2004	1208	1	3.249
##	2589	60949290754	4.041	2004	1208	1	2.948
##	2590	60949341250	3.847	2004	1208	1	2.754
##	2618	67651236094	3.739	2004	1208	1	2.621
##	2624	70450031636	3.739	2004	1208	1	2.621
##	2640	41749085870	5.674	2004	1208	2	4.581
##	2641	52949127270	3.675	2004	1208	2	2.582
##	2645	61149315889	4.521	2004	1208	2	3.403
##	2669	67649687426	3.675	2004	1208	2	2.582
##	2828	34347305000	5.282	2005	1208	1	4.180
##	2865	61249405456	5.282	2005	1208	1	4.180
##	2886	67649786656	4.683	2005	1208	1	3.606
##	3134	70350414060	4.001	2005	1208	2	2.924
##	3135	79957120427	3.708	2005	1208	2	2.631
##	3145	36849063234	5.820	2005	1208	1	4.743
##	3149	60950558911	4.683	2005	1208	1	3.581
##	3166	64949203782	3.688	2005	1208	1	2.611
##	3184	85007789945	4.358	2005	1208	1	3.281
##	3185	85012571734	4.822	2005	1208	1	3.720
##	3186	85014004514	3.822	2005	1208	1	2.745
##	3323	43249162424	3.908	2006	1208	1	2.840
##	3329	60949501532	3.786	2006	1208	1	2.744
##	3337	61149220396	4.126	2006	1208	1	3.084
##	3556	70349380046	4.224	2006	1208	1	3.182
##	3623	34247870091	4.486	2006	1208	1	3.393
##	3637	60950624832	4.486	2006	1208	1	3.444
##	3641	61149487712	4.564	2006	1208	1	3.496
##	3645	67649756861	4.317	2006	1208	1	3.275
##	3646	67649803123	3.908	2006	1208	1	2.840
##	3665	60949428823	4.945	2006	1208	2	3.877
##	3683	70449761232	4.357	2006	1208	2	3.289
##	3784	43249143287	4.082	2007	1208	1	3.080
##	3799	57749162256	4.449	2007	1208	1	3.472
##	3802	57749166879	3.939	2007	1208	1	2.937
##	3818	61149122409	4.555	2007	1208	1	3.578
##	3837	67650088013	4.837	2007	1208	1	3.860
##	3857	60949396166	3.834	2007	1208	2	2.857
##	3888	61949171840	3.629	2007	1208	2	2.652
##	4098	61049174750	4.837	2007	1208	1	3.835
##	4099	61049252344	3.939	2007	1208	1	2.962
##	4100	61049380407	4.335	2007	1208	1	3.307
##	4165	48749117113	4.214	2007	1208	1	3.237
##	4177	61449324554	3.939	2007	1208	1	2.962
##	4186	70350522140	4.654	2007	1208	1	3.677

##	4220	61149608982	3.629	2007	1208	2	2.627
##	4370	60950688715	3.598	2008	1208	1	2.603
##	4406	57749135069	4.149	2008	1208	1	3.179
##	4432	61149730225	4.566	2008	1208	1	3.596
##	4642	45349083255	3.630	2008	1208	2	2.635
##	4697	61449322260	4.300	2008	1208	1	3.305
##	4698	61449425485	4.902	2008	1208	1	3.932
##	4713	79960301345	4.149	2008	1208	1	3.154
##	4722	84977516653	3.598	2008	1208	1	2.628
##	4728	85014203193	3.984	2008	1208	1	3.014
##	4862	75849129999	3.591	2009	1208	1	2.805
##	4981	70350018180	3.591	2009	1208	1	2.780
##	5018	68949127004	4.249	2009	1208	1	3.438
##	5097	65849109286	3.306	2009	1208	1	2.520
##	5099	65849275932	3.837	2009	1208	1	3.026
##	5103	77950760036	3.306	2009	1208	1	2.520
##	5106	84865299965	3.591	2009	1208	1	2.780
##	5107	84873377619	4.425	2009	1208	1	3.639
##	5110	84909053994	3.837	2009	1208	1	3.026
##	5111	84918906073	3.306	2009	1208	1	2.520
##	5117	85008922077	5.433	2009	1208	1	4.647
##	5135	85014598252	3.306	2009	1208	1	2.520
##	5147	77949694173	4.988	2009	1208	2	4.177
##	5371	79957915084	5.520	2010	1208	1	4.693
##	5372	79957915644	3.832	2010	1208	1	3.030
##	5479	77950130289	5.426	2010	1208	1	4.624
##	5480	77950168892	3.586	2010	1208	1	2.759
##	5517	79955439227	3.302	2010	1208	1	2.500
##	5589	84857698666	4.563	2011	1208	2	3.764
##	5598	84855915732	4.594	2011	1208	4	3.820
##	5599	84855916060	4.298	2011	1208	4	3.499
##	5600	84855927603	3.619	2011	1208	4	2.845
##	5728	80051973513	3.887	2011	1208	1	3.088
##	5729	80052003294	3.491	2011	1208	1	2.717
##	5742	84861078242	4.221	2011	1208	1	3.422
##	5816	79958246980	3.887	2011	1208	1	3.088
##	5817	79958252976	3.887	2011	1208	1	3.113
##	5824	80052376495	3.491	2011	1208	1	2.692
##	5866	79960771503	3.887	2011	1208	1	3.113
##	5913	79958063974	3.323	2011	1203	6	2.549
##	5922	79957873219	7.067	2011	1208	1	6.293
##	5994	79952748470	4.221	2011	1208	1	3.447
##	5998	79955375481	4.511	2011	1208	1	3.712
##	6000	79955410959	6.705	2011	1208	1	5.931
##	6007	79956198385	4.972	2011	1208	2	4.173
##	6057	79960875954	3.491	2011	1208	1	2.692
##	6076	84862022685	3.611	2011	1205	2	2.812
##	6166	84870369563	4.382	2012	1208	1	3.534
##	6169	84868627182	3.937	2012	1208	1	3.114
##	6174	84871098188	3.391	2012	1208	1	2.543

```

## 6189 84876955971 4.759 2012      1208      1      3.936
## 6190 84876957904 4.382 2012      1208      1      3.559
## 6191 84876972474 5.865 2012      1208      1      5.042
## 6193 84877007167 5.865 2012      1208      1      5.042
## 6222 84871243321 4.302 2012      1208      2      3.479
## 6330 84868306956 3.686 2012      1208      2      2.863
## 6378 84867794028 4.759 2012      1208      1      3.911
## 6384 84866140479 7.287 2012      1208      2      6.464
## 6386 84868655179 3.559 2012      1208      2      2.736
## 6387 84868667202 3.962 2012      1208      2      3.114
## 6423 84866632349 3.391 2012      1208      1      2.568
## 6425 84866948227 6.455 2012      1208      1      5.632
## 6433 84874479222 7.070 2012      1208      1      6.222
## 6542 84861079400 3.962 2012      1208      2      3.114
## 6544 84864030671 3.567 2012      1208      2      2.769
## 6565 84864673383 3.391 2012      1208      1      2.543
## 6566 84866065375 3.391 2012      1208      1      2.543
## 6587 84859957258 3.520 2012      1202      4      2.697
## 6616 84859326587 7.070 2012      1208      1      6.247
## 6636 84859765129 6.624 2012      1208      1      5.776
## 6655 84872909005 4.721 2012      1202      4      3.898
## 6676 84858641316 3.391 2012      1208      1      2.543
## 6679 84859471937 3.391 2012      1208      1      2.568
## 6681 84860350141 4.382 2012      1208      1      3.559
## 6684 84861541436 3.937 2012      1208      1      3.089
## 6685 84864326825 3.391 2012      1208      1      2.568
## 6689 84856425818 5.835 2012      1208      2      4.987
## 6691 84856507096 3.559 2012      1208      2      2.736
## 6790 84865745109 3.962 2012      1208      2      3.114
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1185 -0.7863 -0.0368  0.6849  6.4636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76931    0.05645   13.63 < 2e-16 ***
## FirstAuthorFemale1 0.02591    0.02535    1.02  0.30674
## Year1997        0.04677    0.07827    0.60  0.55020
## Year1998       -0.07526    0.07137   -1.05  0.29173
## Year1999       -0.09621    0.07152   -1.35  0.17862
## Year2000        0.01574    0.07342    0.21  0.83030
## Year2001        0.10420    0.07851    1.33  0.18450
## Year2002        0.29589    0.07752    3.82  0.00014 ***
## Year2003        0.30350    0.07964    3.81  0.00014 ***

```

```

## Year2004          0.32332      0.07554      4.28  1.9e-05 ***
## Year2005          0.30686      0.07509      4.09  4.4e-05 ***
## Year2006          0.27300      0.07245      3.77  0.00017 ***
## Year2007          0.20736      0.07203      2.88  0.00401 **
## Year2008          0.20020      0.07658      2.61  0.00896 **
## Year2009          0.01698      0.07728      0.22  0.82610
## Year2010          0.03201      0.07925      0.40  0.68629
## Year2011          0.00416      0.07264      0.06  0.95433
## Year2012          0.05407      0.06988      0.77  0.43912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.905
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.019
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
##  16 observations
## c(1850,2216,2654,4345,4560,4645,5023,5089,5252,5254,5420,5453,5460,5613,5628,
## 5675)
## are outliers with |weight| = 0 ( < 1.7e-05);
## 506 weights are ~ = 1. The remaining 5330 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0008 0.8780 0.9300 0.8960 0.9810 0.9990
## Algorithmic parameters:
## tuning.chi          bb          tuning.psi          refine.tol
## 1.55e+00          5.00e-01          4.69e+00          1.00e-07
## rel.tol          solve.tol          eps.outlier          eps.x
## 1.00e-07          1.00e-07          1.71e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01          5.00e-01
## nResample          max.it          best.r.s          k.fast.s          k.max maxit.scale
## 500          50          2          1          1000          200
## trace.lev          mts          compute.rd
## 0          1000          0
## psi          subsampling          cov
## "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.006 1 1.003
## Year 1.006 16 1.000
##
## [1] "List of 154 outliers with residuals above 2.5"
## ScopusId NLCS Year OneField Fields residuals
## 3 33746446640 3.452 1996 1208 1 2.682
## 6 60950433344 4.062 1996 1208 1 3.292
## 133 85015910893 3.345 1996 1208 1 2.575

```

##	240	60949595059	3.564	1997	1208	1	2.747
##	552	61249601308	3.222	1998	1202	4	2.528
##	1080	61249468539	3.822	2000	1208	3	3.037
##	1545	35448994621	3.955	2002	1208	1	2.889
##	1548	60949487473	3.855	2002	1208	1	2.789
##	1588	60950373677	3.920	2002	1208	2	2.854
##	1746	61249391353	4.222	2002	1208	1	3.156
##	1751	61249678334	3.746	2002	1208	1	2.655
##	1785	48949111059	4.224	2002	1208	2	3.133
##	1793	61949171102	3.665	2002	1208	2	2.599
##	1851	60949941202	3.850	2002	1202	3	2.784
##	1994	46449113576	4.314	2003	1208	1	3.241
##	1999	60949679561	3.954	2003	1208	1	2.856
##	2024	63449141239	3.780	2003	1208	1	2.707
##	2027	27844509137	4.429	2003	1208	2	3.356
##	2028	33645366634	4.044	2003	1208	2	2.946
##	2181	60950014779	3.580	2003	1208	1	2.507
##	2203	64249135095	4.034	2003	1208	1	2.961
##	2220	33749432390	3.807	2003	1208	2	2.683
##	2223	34248348215	5.357	2003	1208	2	4.284
##	2378	50549091357	4.213	2004	1208	1	3.120
##	2389	60950006630	3.947	2004	1208	1	2.854
##	2392	60950365720	3.622	2004	1208	1	2.529
##	2395	60950504818	4.213	2004	1208	1	3.120
##	2398	60950612949	3.947	2004	1208	1	2.854
##	2400	61049193685	3.622	2004	1208	1	2.504
##	2422	19644393451	3.827	2004	1208	2	2.734
##	2444	60950717875	3.827	2004	1208	2	2.734
##	2587	49849105643	4.367	2004	1208	1	3.249
##	2589	60949290754	4.041	2004	1208	1	2.948
##	2590	60949341250	3.847	2004	1208	1	2.754
##	2618	67651236094	3.739	2004	1208	1	2.621
##	2624	70450031636	3.739	2004	1208	1	2.621
##	2640	41749085870	5.674	2004	1208	2	4.581
##	2641	52949127270	3.675	2004	1208	2	2.582
##	2645	61149315889	4.521	2004	1208	2	3.403
##	2669	67649687426	3.675	2004	1208	2	2.582
##	2828	34347305000	5.282	2005	1208	1	4.180
##	2865	61249405456	5.282	2005	1208	1	4.180
##	2886	67649786656	4.683	2005	1208	1	3.606
##	3134	70350414060	4.001	2005	1208	2	2.924
##	3135	79957120427	3.708	2005	1208	2	2.631
##	3145	36849063234	5.820	2005	1208	1	4.743
##	3149	60950558911	4.683	2005	1208	1	3.581
##	3166	64949203782	3.688	2005	1208	1	2.611
##	3184	85007789945	4.358	2005	1208	1	3.281
##	3185	85012571734	4.822	2005	1208	1	3.720
##	3186	85014004514	3.822	2005	1208	1	2.745
##	3323	43249162424	3.908	2006	1208	1	2.840
##	3329	60949501532	3.786	2006	1208	1	2.744

##	3337	61149220396	4.126	2006	1208	1	3.084
##	3556	70349380046	4.224	2006	1208	1	3.182
##	3623	34247870091	4.486	2006	1208	1	3.393
##	3637	60950624832	4.486	2006	1208	1	3.444
##	3641	61149487712	4.564	2006	1208	1	3.496
##	3645	67649756861	4.317	2006	1208	1	3.275
##	3646	67649803123	3.908	2006	1208	1	2.840
##	3665	60949428823	4.945	2006	1208	2	3.877
##	3683	70449761232	4.357	2006	1208	2	3.289
##	3784	43249143287	4.082	2007	1208	1	3.080
##	3799	57749162256	4.449	2007	1208	1	3.472
##	3802	57749166879	3.939	2007	1208	1	2.937
##	3818	61149122409	4.555	2007	1208	1	3.578
##	3837	67650088013	4.837	2007	1208	1	3.860
##	3857	60949396166	3.834	2007	1208	2	2.857
##	3888	61949171840	3.629	2007	1208	2	2.652
##	4098	61049174750	4.837	2007	1208	1	3.835
##	4099	61049252344	3.939	2007	1208	1	2.962
##	4100	61049380407	4.335	2007	1208	1	3.307
##	4165	48749117113	4.214	2007	1208	1	3.237
##	4177	61449324554	3.939	2007	1208	1	2.962
##	4186	70350522140	4.654	2007	1208	1	3.677
##	4220	61149608982	3.629	2007	1208	2	2.627
##	4370	60950688715	3.598	2008	1208	1	2.603
##	4406	57749135069	4.149	2008	1208	1	3.179
##	4432	61149730225	4.566	2008	1208	1	3.596
##	4642	45349083255	3.630	2008	1208	2	2.635
##	4697	61449322260	4.300	2008	1208	1	3.305
##	4698	61449425485	4.902	2008	1208	1	3.932
##	4713	79960301345	4.149	2008	1208	1	3.154
##	4722	84977516653	3.598	2008	1208	1	2.628
##	4728	85014203193	3.984	2008	1208	1	3.014
##	4862	75849129999	3.591	2009	1208	1	2.805
##	4981	70350018180	3.591	2009	1208	1	2.780
##	5018	68949127004	4.249	2009	1208	1	3.438
##	5097	65849109286	3.306	2009	1208	1	2.520
##	5099	65849275932	3.837	2009	1208	1	3.026
##	5103	77950760036	3.306	2009	1208	1	2.520
##	5106	84865299965	3.591	2009	1208	1	2.780
##	5107	84873377619	4.425	2009	1208	1	3.639
##	5110	84909053994	3.837	2009	1208	1	3.026
##	5111	84918906073	3.306	2009	1208	1	2.520
##	5117	85008922077	5.433	2009	1208	1	4.647
##	5135	85014598252	3.306	2009	1208	1	2.520
##	5147	77949694173	4.988	2009	1208	2	4.177
##	5371	79957915084	5.520	2010	1208	1	4.693
##	5372	79957915644	3.832	2010	1208	1	3.030
##	5479	77950130289	5.426	2010	1208	1	4.624
##	5480	77950168892	3.586	2010	1208	1	2.759
##	5517	79955439227	3.302	2010	1208	1	2.500



##	5589	84857698666	4.563	2011	1208	2	3.764
##	5598	84855915732	4.594	2011	1208	4	3.820
##	5599	84855916060	4.298	2011	1208	4	3.499
##	5600	84855927603	3.619	2011	1208	4	2.845
##	5728	80051973513	3.887	2011	1208	1	3.088
##	5729	80052003294	3.491	2011	1208	1	2.717
##	5742	84861078242	4.221	2011	1208	1	3.422
##	5816	79958246980	3.887	2011	1208	1	3.088
##	5817	79958252976	3.887	2011	1208	1	3.113
##	5824	80052376495	3.491	2011	1208	1	2.692
##	5866	79960771503	3.887	2011	1208	1	3.113
##	5913	79958063974	3.323	2011	1203	6	2.549
##	5922	79957873219	7.067	2011	1208	1	6.293
##	5994	79952748470	4.221	2011	1208	1	3.447
##	5998	79955375481	4.511	2011	1208	1	3.712
##	6000	79955410959	6.705	2011	1208	1	5.931
##	6007	79956198385	4.972	2011	1208	2	4.173
##	6057	79960875954	3.491	2011	1208	1	2.692
##	6076	84862022685	3.611	2011	1205	2	2.812
##	6166	84870369563	4.382	2012	1208	1	3.534
##	6169	84868627182	3.937	2012	1208	1	3.114
##	6174	84871098188	3.391	2012	1208	1	2.543
##	6189	84876955971	4.759	2012	1208	1	3.936
##	6190	84876957904	4.382	2012	1208	1	3.559
##	6191	84876972474	5.865	2012	1208	1	5.042
##	6193	84877007167	5.865	2012	1208	1	5.042
##	6222	84871243321	4.302	2012	1208	2	3.479
##	6330	84868306956	3.686	2012	1208	2	2.863
##	6378	84867794028	4.759	2012	1208	1	3.911
##	6384	84866140479	7.287	2012	1208	2	6.464
##	6386	84868655179	3.559	2012	1208	2	2.736
##	6387	84868667202	3.962	2012	1208	2	3.114
##	6423	84866632349	3.391	2012	1208	1	2.568
##	6425	84866948227	6.455	2012	1208	1	5.632
##	6433	84874479222	7.070	2012	1208	1	6.222
##	6542	84861079400	3.962	2012	1208	2	3.114
##	6544	84864030671	3.567	2012	1208	2	2.769
##	6565	84864673383	3.391	2012	1208	1	2.543
##	6566	84866065375	3.391	2012	1208	1	2.543
##	6587	84859957258	3.520	2012	1202	4	2.697
##	6616	84859326587	7.070	2012	1208	1	6.247
##	6636	84859765129	6.624	2012	1208	1	5.776
##	6655	84872909005	4.721	2012	1202	4	3.898
##	6676	84858641316	3.391	2012	1208	1	2.543
##	6679	84859471937	3.391	2012	1208	1	2.568
##	6681	84860350141	4.382	2012	1208	1	3.559
##	6684	84861541436	3.937	2012	1208	1	3.089
##	6685	84864326825	3.391	2012	1208	1	2.568
##	6689	84856425818	5.835	2012	1208	2	4.987
##	6691	84856507096	3.559	2012	1208	2	2.736

```

## 6790 84865745109 3.962 2012      1208      2      3.114
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1170 -0.7878 -0.0378  0.6850  6.4619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77035    0.05649   13.64 < 2e-16 ***
## LastAuthorFemale1 0.02310    0.02533    0.91  0.36190
## Year1997        0.04642    0.07827    0.59  0.55316
## Year1998       -0.07547    0.07138   -1.06  0.29039
## Year1999       -0.09620    0.07150   -1.35  0.17856
## Year2000        0.01563    0.07344    0.21  0.83151
## Year2001        0.10480    0.07852    1.33  0.18205
## Year2002        0.29609    0.07753    3.82  0.00014 ***
## Year2003        0.30371    0.07966    3.81  0.00014 ***
## Year2004        0.32357    0.07557    4.28  1.9e-05 ***
## Year2005        0.30682    0.07509    4.09  4.4e-05 ***
## Year2006        0.27327    0.07247    3.77  0.00016 ***
## Year2007        0.20756    0.07204    2.88  0.00398 **
## Year2008        0.20014    0.07660    2.61  0.00900 **
## Year2009        0.01740    0.07729    0.23  0.82188
## Year2010        0.03225    0.07924    0.41  0.68401
## Year2011        0.00459    0.07263    0.06  0.94967
## Year2012        0.05473    0.06988    0.78  0.43356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.906
## Multiple R-squared:  0.0218, Adjusted R-squared:  0.019
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 16 observations
c(1850,2216,2654,4345,4560,4645,5023,5089,5252,5254,5420,5453,5460,5613,5628,
5675)
## are outliers with |weight| = 0 ( < 1.7e-05);
## 506 weights are ~= 1. The remaining 5330 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0008 0.8780 0.9310 0.8970 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x

```

```

##          1.00e-07          1.00e-07          1.71e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##          500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5852"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1209"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    7    3    3    3    9    8    5   15   14   32   28   22   19   29
## 2011 2012
##   15   13
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    7    2    3    3    5    7    4   14   13   29   24   20   13   27
## 2011 2012
##   13   11
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1    7    2    3    3    5    7    4   14   13   28   24   20   13   27
## 2011 2012
##   13   10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.47915977772535"
## [1] "Male first author team size 2018 geometric mean: 1.09372354772644"

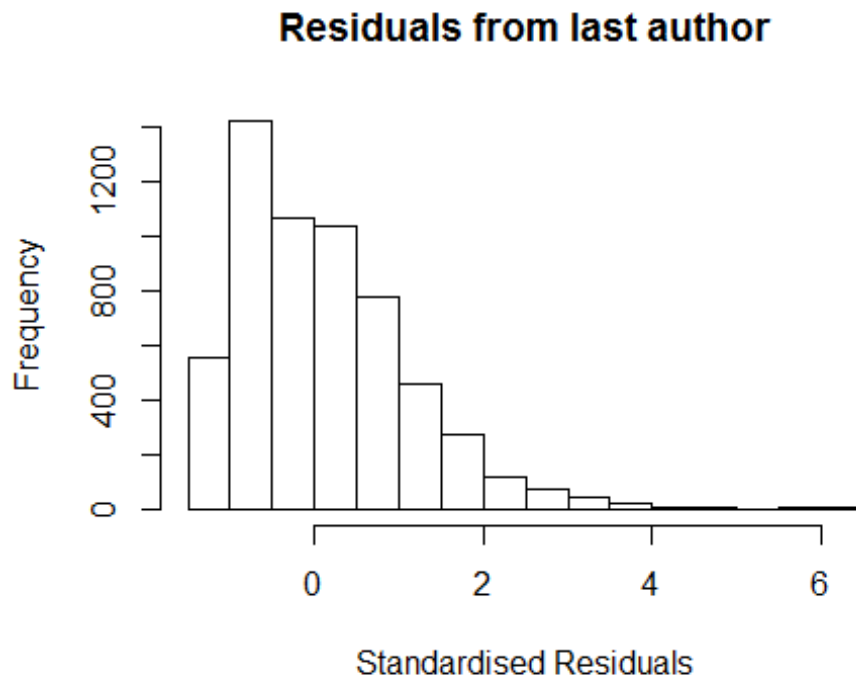
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##

```

```
## [1] "Female last author team size 2018 geometric mean: 1.23284673944207"
## [1] "Male last author team size 2018 geometric mean: 1.24120023026734"

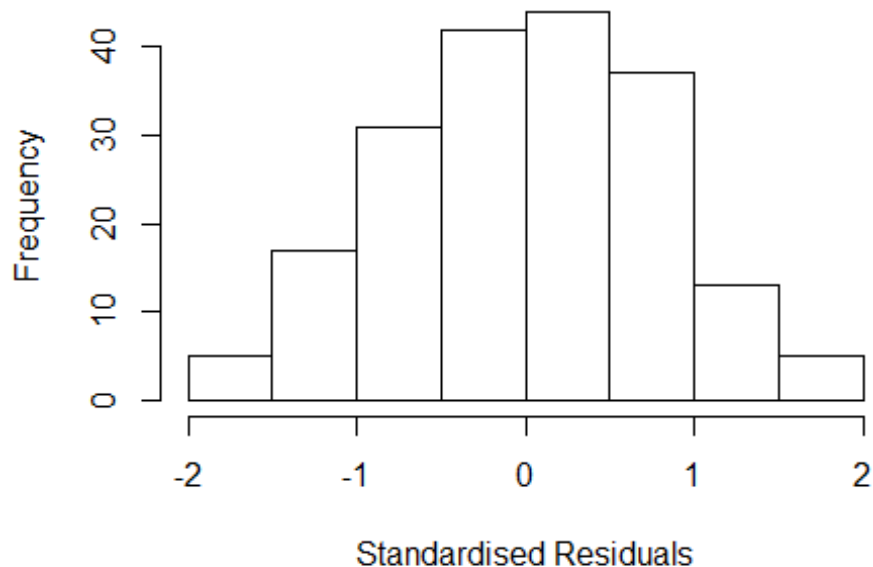
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	5.762	1	2.400
LastAuthorFemale	9.337	1	3.056
UniqueAuthors	7.429	3	1.397
Year	24.522	16	1.105

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7178 -0.5847 0.0499 0.6469 1.7623
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.69e-15 2.47e-07 0.00 1.00000
## FirstAuthorFemale1 4.21e-02 2.02e-01 0.21 0.83485
## LastAuthorFemale1 -8.55e-02 2.13e-01 -0.40 0.68786
## UniqueAuthors2 3.15e-01 1.96e-01 1.60 0.11114
## UniqueAuthors3 4.60e-01 3.82e-01 1.20 0.23076
## UniqueAuthors4 5.29e-01 1.96e-01 2.70 0.00769 **
## Year1997 4.62e-01 1.28e-01 3.59 0.00043 ***
## Year1998 5.94e-01 4.24e-01 1.40 0.16247
## Year1999 1.38e+00 1.36e-01 10.08 < 2e-16 ***
## Year2000 6.65e-01 3.37e-01 1.97 0.04993 *
```

```

## Year2001      8.36e-01  2.98e-01  2.80  0.00564 **
## Year2002      1.72e+00  4.01e-01  4.29  3.0e-05 ***
## Year2003      1.84e+00  4.60e-01  4.00  9.2e-05 ***
## Year2004      1.43e+00  2.29e-01  6.22  3.6e-09 ***
## Year2005      1.37e+00  2.78e-01  4.94  1.9e-06 ***
## Year2006      1.52e+00  2.14e-01  7.10  3.1e-11 ***
## Year2007      8.71e-01  1.99e-01  4.37  2.2e-05 ***
## Year2008      1.07e+00  1.50e-01  7.12  2.8e-11 ***
## Year2009      1.03e+00  2.87e-01  3.57  0.00046 ***
## Year2010      1.12e+00  2.00e-01  5.61  8.1e-08 ***
## Year2011      1.33e+00  2.57e-01  5.15  6.9e-07 ***
## Year2012      1.14e+00  2.47e-01  4.60  8.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.911
## Multiple R-squared:  0.151, Adjusted R-squared:  0.0477
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 173 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.688  0.897  0.952  0.926  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.936 1      2.222
## LastAuthorFemale  5.297 1      2.302
## Year              2.655 16      1.031
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##

```

```

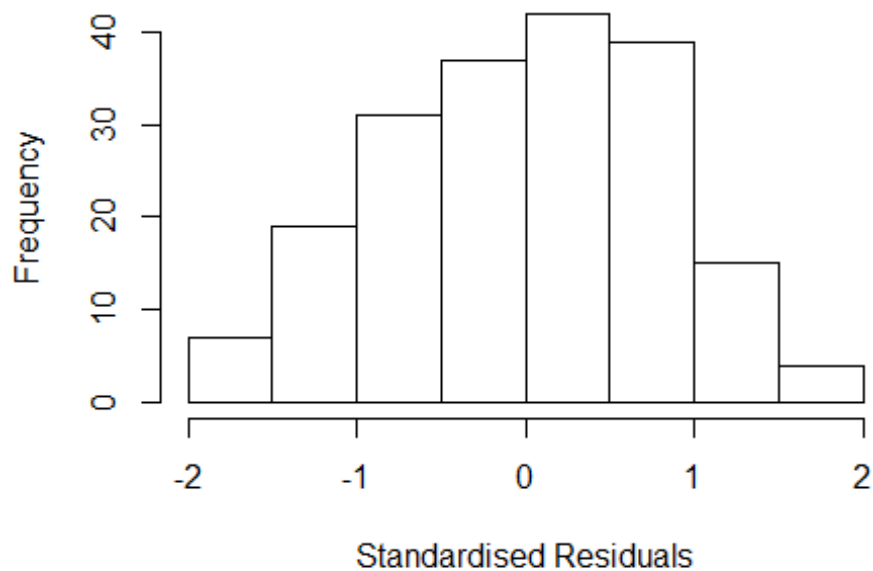
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8569 -0.5963  0.0149  0.6195  1.9975
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.48e-17   2.97e-07    0.00  1.00000
## FirstAuthorFemale1  5.28e-02   2.05e-01    0.26  0.79700
## LastAuthorFemale1 -1.02e-01   2.09e-01   -0.49  0.62503
## Year1997        4.62e-01   1.29e-01    3.59  0.00043 ***
## Year1998        5.97e-01   4.22e-01    1.42  0.15847
## Year1999        1.56e+00   2.29e-01    6.79  1.7e-10 ***
## Year2000        6.67e-01   3.36e-01    1.99  0.04853 *
## Year2001        9.52e-01   3.16e-01    3.01  0.00296 **
## Year2002        1.86e+00   3.32e-01    5.59  8.4e-08 ***
## Year2003        1.84e+00   4.62e-01    3.99  9.7e-05 ***
## Year2004        1.46e+00   2.23e-01    6.53  7.0e-10 ***
## Year2005        1.43e+00   2.82e-01    5.08  9.6e-07 ***
## Year2006        1.59e+00   2.10e-01    7.58  1.9e-12 ***
## Year2007        8.99e-01   2.06e-01    4.36  2.2e-05 ***
## Year2008        1.14e+00   1.46e-01    7.78  6.1e-13 ***
## Year2009        1.06e+00   2.80e-01    3.78  0.00021 ***
## Year2010        1.21e+00   1.97e-01    6.13  5.6e-09 ***
## Year2011        1.39e+00   2.48e-01    5.62  7.6e-08 ***
## Year2012        1.14e+00   2.47e-01    4.61  7.7e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.902
## Multiple R-squared:  0.134, Adjusted R-squared:  0.0446
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 176 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.603  0.889  0.952  0.924  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      5.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2          1          1000      200

```

```
##      trace.lev      mts  compute.rd
##           0      1000      0
##           psi      subsampling      cov
##           "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```

### Residuals from first and last author



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.019 1      1.421
## Year              2.019 16     1.022

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
```

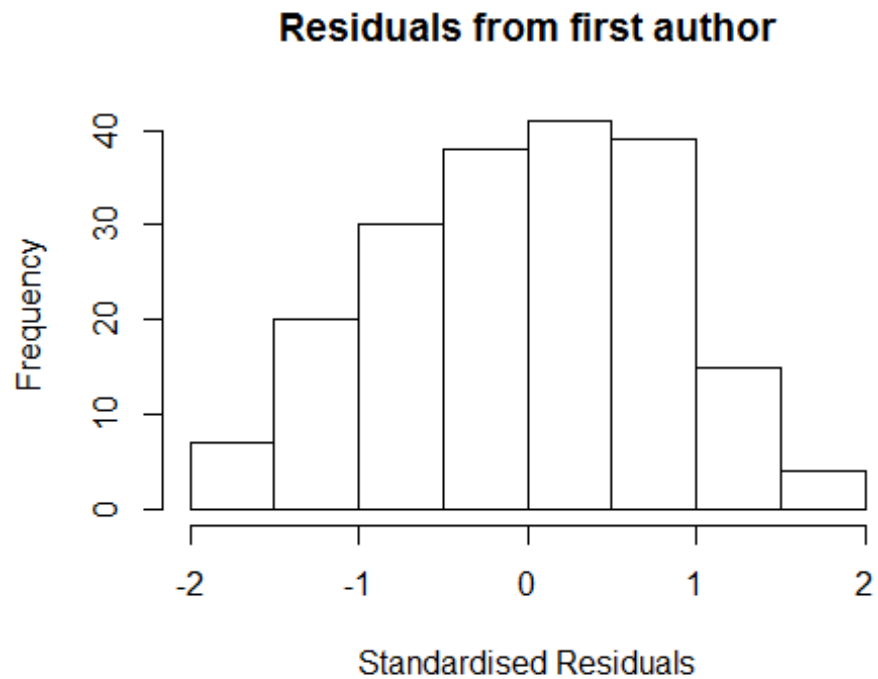


```

## -1.8584 -0.5958 0.0171 0.6079 1.9887
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.57e-15  0.00e+00   -Inf    < 2e-16 ***
## FirstAuthorFemale1 -3.46e-02  1.32e-01   -0.26  0.79430
## Year1997       4.60e-01  1.28e-01    3.60  0.00041 ***
## Year1998       5.90e-01  4.28e-01    1.38  0.17009
## Year1999       1.52e+00  1.96e-01    7.76  6.7e-13 ***
## Year2000       6.62e-01  3.40e-01    1.95  0.05283 .
## Year2001       9.48e-01  3.18e-01    2.99  0.00324 **
## Year2002       1.86e+00  3.32e-01    5.60  8.0e-08 ***
## Year2003       1.84e+00  4.64e-01    3.96  0.00011 ***
## Year2004       1.44e+00  2.17e-01    6.63  4.0e-10 ***
## Year2005       1.44e+00  2.82e-01    5.10  8.7e-07 ***
## Year2006       1.58e+00  2.05e-01    7.72  8.6e-13 ***
## Year2007       8.94e-01  2.06e-01    4.33  2.5e-05 ***
## Year2008       1.13e+00  1.44e-01    7.84  4.1e-13 ***
## Year2009       1.06e+00  2.82e-01    3.76  0.00023 ***
## Year2010       1.20e+00  1.95e-01    6.16  4.8e-09 ***
## Year2011       1.39e+00  2.52e-01    5.54  1.1e-07 ***
## Year2012       1.13e+00  2.45e-01    4.62  7.5e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.892
## Multiple R-squared:  0.133, Adjusted R-squared:  0.0497
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 177 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.598  0.885   0.952   0.922   0.982   0.999
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.15e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"

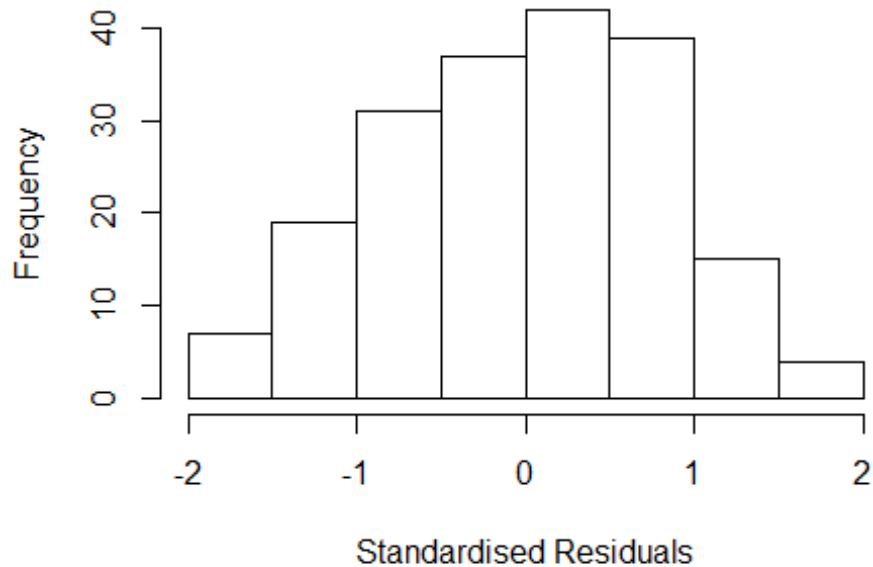
```

```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```



##		GVIF	Df	GVIF^(1/(2*Df))
##	LastAuthorFemale	2.18	1	1.477
##	Year	2.18	16	1.025

## Residuals from last author



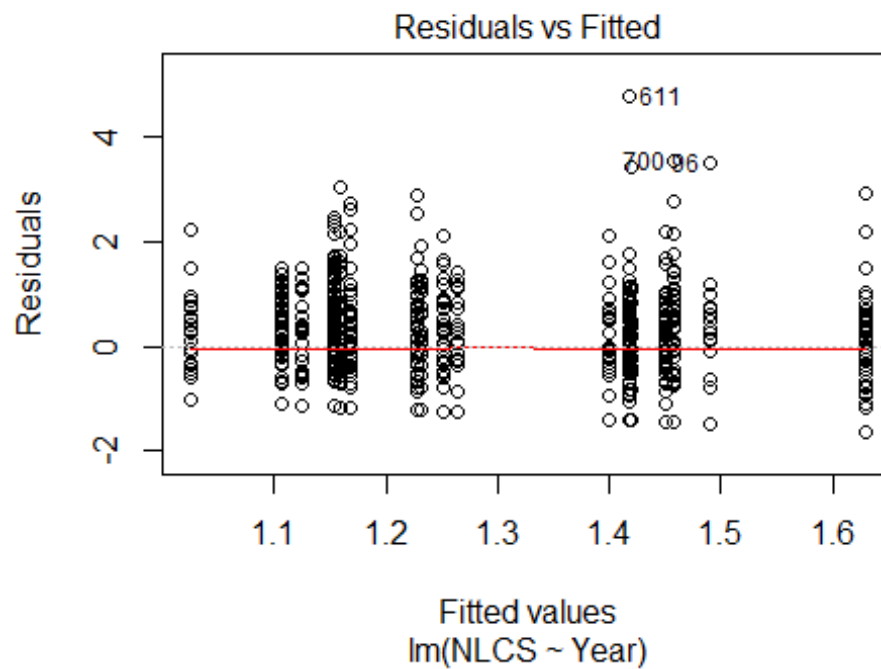
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.85796 -0.60197 0.00496 0.61276 1.99978
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 7.01e-15 0.00e+00 Inf < 2e-16 ***
## LastAuthorFemale1 -5.70e-02 1.34e-01 -0.43 0.67083
## Year1997 4.63e-01 1.29e-01 3.58 0.00044 ***
## Year1998 6.01e-01 4.18e-01 1.44 0.15260
## Year1999 1.54e+00 2.18e-01 7.08 3.3e-11 ***
## Year2000 6.70e-01 3.34e-01 2.00 0.04655 *
## Year2001 9.53e-01 3.15e-01 3.02 0.00288 **
## Year2002 1.86e+00 3.32e-01 5.60 8.1e-08 ***
## Year2003 1.85e+00 4.62e-01 4.00 9.4e-05 ***
## Year2004 1.46e+00 2.23e-01 6.52 7.1e-10 ***
## Year2005 1.44e+00 2.78e-01 5.17 6.4e-07 ***
## Year2006 1.60e+00 2.12e-01 7.54 2.4e-12 ***
```

```

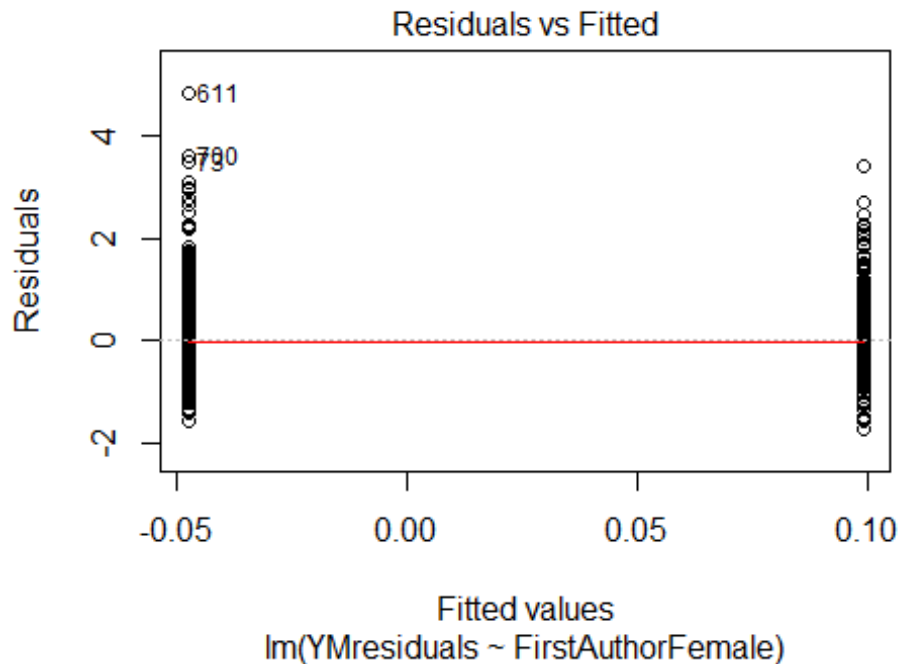
## Year2007          9.01e-01    2.06e-01    4.37  2.1e-05 ***
## Year2008          1.14e+00    1.45e-01    7.86  3.8e-13 ***
## Year2009          1.07e+00    2.72e-01    3.93  0.00012 ***
## Year2010          1.21e+00    1.97e-01    6.14  5.4e-09 ***
## Year2011          1.40e+00    2.48e-01    5.66  5.9e-08 ***
## Year2012          1.15e+00    2.46e-01    4.66  6.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.895
## Multiple R-squared:  0.134, Adjusted R-squared:  0.0502
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 177 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.597  0.887  0.951  0.923  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 194"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1210"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   43   37   29   40   45   43   58   68   36   61   66   97   81   59   95
## 2011 2012
##  109  165
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   38   31   26   33   41   33   51   49   31   59   59   79   68   52   80
## 2011 2012

```

```
## 100 140
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 31 26 33 40 31 49 47 28 56 55 79 67 49 78
## 2011 2012
## 95 138
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.06
```

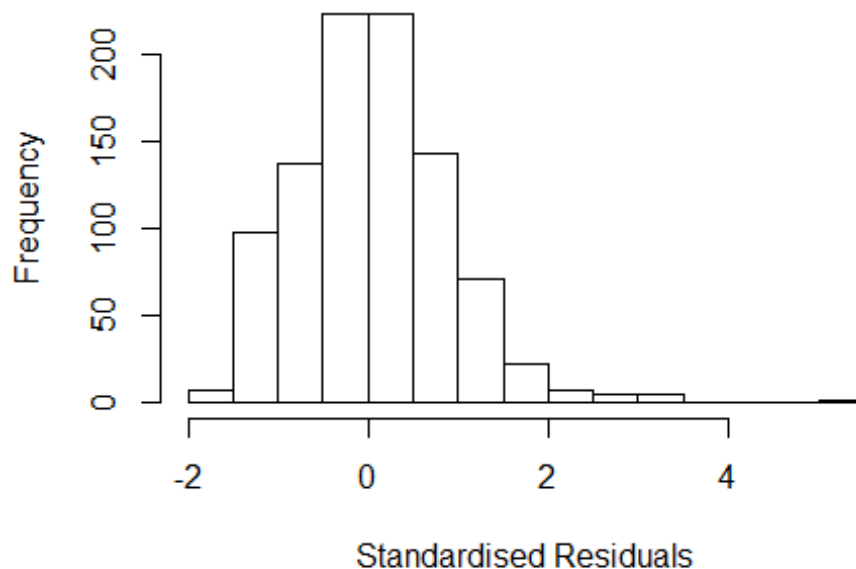


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.17, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 1.41425556191587"
## [1] "Male first author team size 2018 geometric mean: 1.4097352867638"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.33717019954735"
## [1] "Male last author team size 2018 geometric mean: 1.45158344508673"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.330 1      1.527
## LastAuthorFemale 2.312 1      1.520
## UniqueAuthors    1.330 4      1.036
## Year              1.486 16     1.012
```

## Residuals from first and last author and team size



```
## [1] "List of 10 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 73      0031119324 4.853 1997      1210      2      3.277
## 96      0000585627 4.987 1998      1210      1      3.156
## 276     0036565284 4.562 2002      1210      2      2.730
## 432     62849128009 3.778 2005      1210      1      2.630
## 437     33745181909 4.105 2005      1210      2      3.130
## 508     33748493765 4.219 2006      1210      2      3.279
## 611     36048982830 6.200 2007      1210      2      5.007
## 700     57049098170 5.016 2008      1210      2      3.311
## 835     78249271418 3.777 2010      1210      2      2.913
## 1047    84871110200 3.607 2012      1210      2      2.672
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7658 -0.5332  0.0133  0.5305  5.0067
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1140     0.1574   7.08 2.9e-12 ***
```

```

## FirstAuthorFemale1  0.1116      0.0875      1.28      0.2024
## LastAuthorFemale1   0.0609      0.0883      0.69      0.4905
## UniqueAuthors2      0.4593      0.0723      6.35      3.3e-10 ***
## UniqueAuthors3      0.3962      0.1463      2.71      0.0069 **
## UniqueAuthors4      0.6870      0.1704      4.03      6.0e-05 ***
## UniqueAuthors5      0.3693      0.2533      1.46      0.1451
## Year1997             0.0661      0.2160      0.31      0.7597
## Year1998             0.1486      0.2452      0.61      0.5445
## Year1999            -0.0221      0.2022     -0.11      0.9131
## Year2000            -0.2000      0.1993     -1.00      0.3157
## Year2001            -0.2778      0.1865     -1.49      0.1367
## Year2002             0.2586      0.1983      1.30      0.1925
## Year2003            -0.0375      0.2055     -0.18      0.8554
## Year2004             0.1299      0.2275      0.57      0.5682
## Year2005            -0.1388      0.1955     -0.71      0.4777
## Year2006            -0.1744      0.2057     -0.85      0.3967
## Year2007             0.0793      0.1754      0.45      0.6514
## Year2008             0.1316      0.2005      0.66      0.5120
## Year2009             0.0576      0.1960      0.29      0.7688
## Year2010            -0.2498      0.1770     -1.41      0.1586
## Year2011            -0.1732      0.1730     -1.00      0.3169
## Year2012            -0.1789      0.1761     -1.02      0.3099
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.816
## Multiple R-squared:  0.0983, Adjusted R-squared:  0.0766
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 479 is an outlier with |weight| = 0 ( < 0.00011);
## 83 weights are ~= 1. The remaining 856 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0623 0.8690 0.9520 0.9060 0.9880 0.9990
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00             5.00e-01           4.69e+00           1.00e-07
##           rel.tol             solve.tol           eps.outlier           eps.x
##           1.00e-07             1.00e-07           1.06e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01             5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi             subsampling             cov
##           "bisquare"             "nonsingular"             ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)

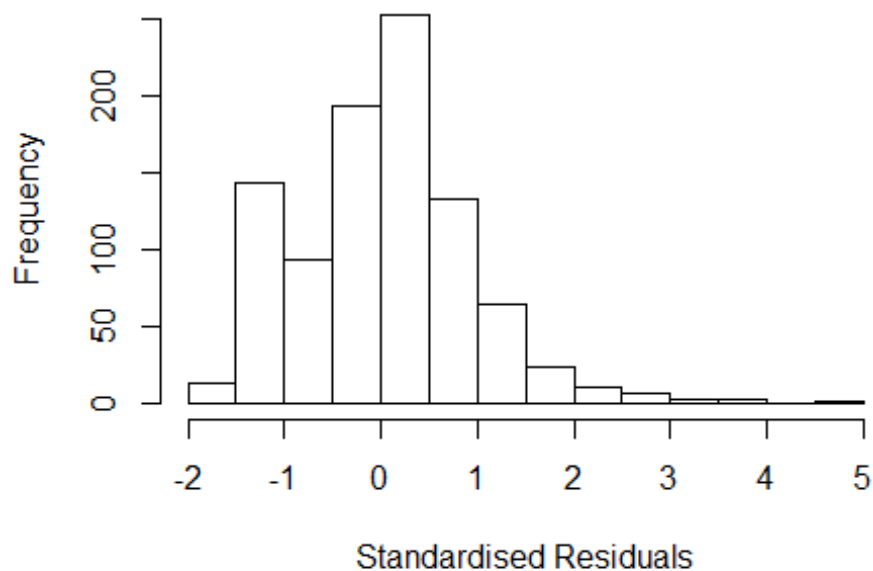
```



```
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
```

```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.391  1      1.546
## LastAuthorFemale  2.342  1      1.531
## Year              1.136 16      1.004
```

### Residuals from first and last author



```
## [1] "List of 12 outliers with residuals above 2.5"
```

	ScopusId	NLCS	Year	OneField	Fields	residuals
## 73	0031119324	4.853	1997	1210	2	3.554
## 96	0000585627	4.987	1998	1210	1	3.473
## 276	0036565284	4.562	2002	1210	2	3.058
## 432	62849128009	3.778	2005	1210	1	2.508
## 437	33745181909	4.105	2005	1210	2	2.996
## 508	33748493765	4.219	2006	1210	2	3.188
## 611	36048982830	6.200	2007	1210	2	4.897
## 700	57049098170	5.016	2008	1210	2	3.676
## 703	42349108299	4.239	2008	1210	1	2.738
## 811	79955423767	3.903	2010	1210	2	2.877
## 835	78249271418	3.777	2010	1210	2	2.751
## 1047	84871110200	3.607	2012	1210	2	2.570

```
##
```

```
## Call:
```

```
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered,
```

```
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
```

```
## \--> method = "MM"
```

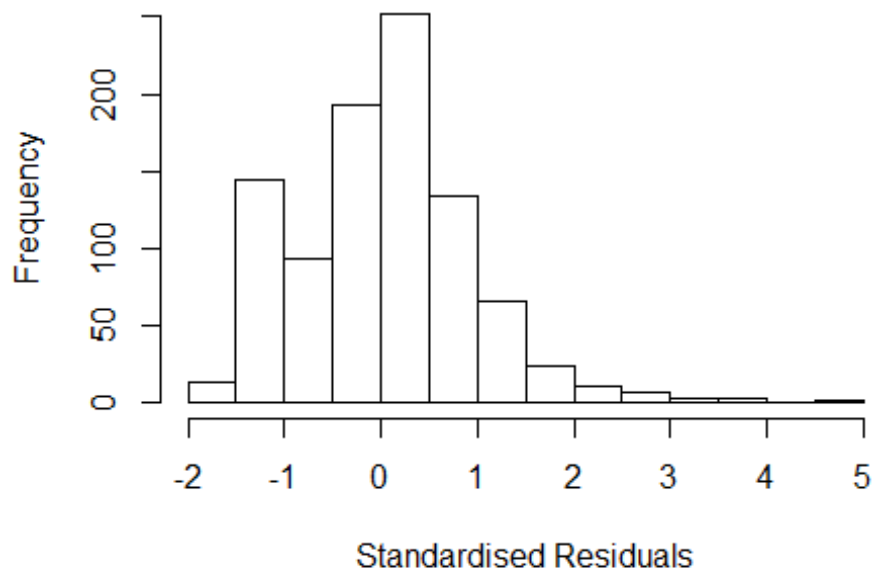
```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6647 -0.5537  0.0432  0.5347  4.8971
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19017    0.16585    7.18 1.5e-12 ***
## FirstAuthorFemale1 0.19255    0.09138    2.11  0.035 *
## LastAuthorFemale1 -0.03164    0.09248   -0.34  0.732
## Year1997         0.10854    0.22179    0.49  0.625
## Year1998         0.16270    0.25004    0.65  0.515
## Year1999         0.02584    0.21099    0.12  0.903
## Year2000        -0.17356    0.20579   -0.84  0.399
## Year2001        -0.30270    0.19715   -1.54  0.125
## Year2002         0.31363    0.20866    1.50  0.133
## Year2003         0.00771    0.21260    0.04  0.971
## Year2004         0.10820    0.23558    0.46  0.646
## Year2005        -0.08097    0.20558   -0.39  0.694
## Year2006        -0.15871    0.21310   -0.74  0.457
## Year2007         0.11268    0.18722    0.60  0.547
## Year2008         0.14952    0.20833    0.72  0.473
## Year2009         0.15469    0.20470    0.76  0.450
## Year2010        -0.16404    0.18415   -0.89  0.373
## Year2011        -0.15184    0.18334   -0.83  0.408
## Year2012        -0.15301    0.18559   -0.82  0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.831
## Multiple R-squared:  0.0426, Adjusted R-squared:  0.0238
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 479 is an outlier with |weight| = 0 ( < 0.00011);
## 82 weights are ~ 1. The remaining 857 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0118 0.8630 0.9530 0.9040 0.9880 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"

```

```
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.1 1      1.049
## Year              1.1 16     1.003
```

### Residuals from first author



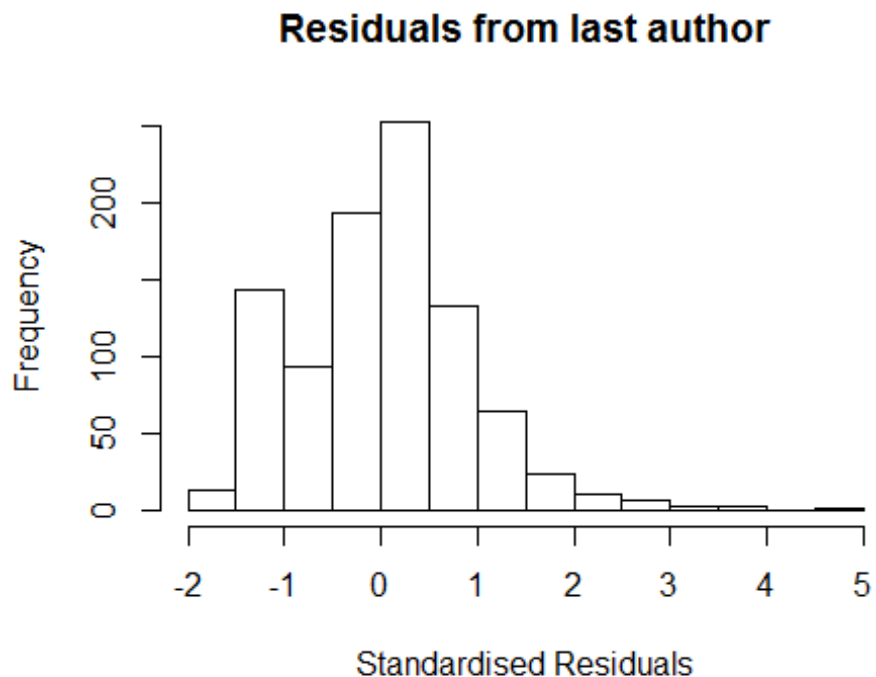
```
## [1] "List of 12 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 73      0031119324 4.853 1997    1210      2      3.554
## 96      0000585627 4.987 1998    1210      1      3.473
## 276     0036565284 4.562 2002    1210      2      3.058
## 432     62849128009 3.778 2005    1210      1      2.508
## 437     33745181909 4.105 2005    1210      2      2.996
## 508     33748493765 4.219 2006    1210      2      3.188
## 611     36048982830 6.200 2007    1210      2      4.897
## 700     57049098170 5.016 2008    1210      2      3.676
## 703     42349108299 4.239 2008    1210      1      2.738
## 811     79955423767 3.903 2010    1210      2      2.877
## 835     78249271418 3.777 2010    1210      2      2.751
## 1047    84871110200 3.607 2012    1210      2      2.570
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
```

```

## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6697 -0.5521  0.0447  0.5363  4.8977
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18839    0.16577   7.17 1.6e-12 ***
## FirstAuthorFemale1 0.16819    0.06215   2.71  0.0069 **
## Year1997        0.10992    0.22181   0.50  0.6203
## Year1998        0.16160    0.25013   0.65  0.5184
## Year1999        0.02461    0.21098   0.12  0.9072
## Year2000       -0.17379    0.20589  -0.84  0.3988
## Year2001       -0.30258    0.19708  -1.54  0.1251
## Year2002        0.31314    0.20842   1.50  0.1333
## Year2003        0.00725    0.21253   0.03  0.9728
## Year2004        0.10479    0.23518   0.45  0.6560
## Year2005       -0.08085    0.20582  -0.39  0.6946
## Year2006       -0.16000    0.21297  -0.75  0.4527
## Year2007        0.11396    0.18729   0.61  0.5430
## Year2008        0.15073    0.20849   0.72  0.4699
## Year2009        0.15488    0.20476   0.76  0.4496
## Year2010       -0.16430    0.18421  -0.89  0.3727
## Year2011       -0.15062    0.18343  -0.82  0.4118
## Year2012       -0.15230    0.18565  -0.82  0.4122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.83
## Multiple R-squared:  0.0425, Adjusted R-squared:  0.0248
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 479 is an outlier with |weight| = 0 ( < 0.00011);
## 81 weights are ~ 1. The remaining 858 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0111 0.8630 0.9530 0.9030 0.9880 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"

```

```
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.081 1      1.040
## Year             1.081 16      1.002
```



```
## [1] "List of 12 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 73    0031119324 4.853 1997    1210      2    3.554
## 96    0000585627 4.987 1998    1210      1    3.473
## 276   0036565284 4.562 2002    1210      2    3.058
## 432   62849128009 3.778 2005    1210      1    2.508
## 437   33745181909 4.105 2005    1210      2    2.996
## 508   33748493765 4.219 2006    1210      2    3.188
## 611   36048982830 6.200 2007    1210      2    4.897
## 700   57049098170 5.016 2008    1210      2    3.676
## 703   42349108299 4.239 2008    1210      1    2.738
## 811   79955423767 3.903 2010    1210      2    2.877
## 835   78249271418 3.777 2010    1210      2    2.751
## 1047  84871110200 3.607 2012    1210      2    2.570
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
```

```

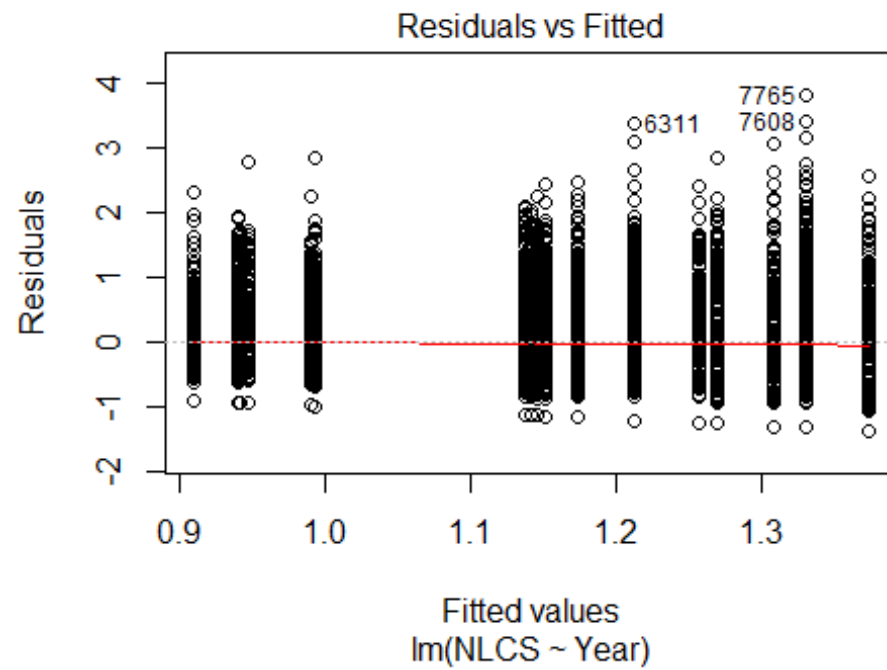
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.642 -0.547  0.038  0.539  4.878
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19601    0.16761    7.14 1.9e-12 ***
## LastAuthorFemale1 0.12222    0.06255    1.95  0.051 .
## Year1997        0.11785    0.22345    0.53  0.598
## Year1998        0.16411    0.25243    0.65  0.516
## Year1999        0.03122    0.21143    0.15  0.883
## Year2000       -0.15917    0.20635   -0.77  0.441
## Year2001       -0.30957    0.19790   -1.56  0.118
## Year2002        0.32377    0.20860    1.55  0.121
## Year2003        0.00595    0.21426    0.03  0.978
## Year2004        0.10173    0.23490    0.43  0.665
## Year2005       -0.07346    0.20685   -0.36  0.723
## Year2006       -0.15861    0.21326   -0.74  0.457
## Year2007        0.12604    0.18853    0.67  0.504
## Year2008        0.16454    0.21009    0.78  0.434
## Year2009        0.17158    0.20449    0.84  0.402
## Year2010       -0.15525    0.18567   -0.84  0.403
## Year2011       -0.13104    0.18446   -0.71  0.478
## Year2012       -0.13439    0.18689   -0.72  0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.832
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.0205
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 479 is an outlier with |weight| = 0 ( < 0.00011);
## 81 weights are ~ 1. The remaining 858 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0145 0.8570 0.9540 0.9040 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.06e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"

```

```

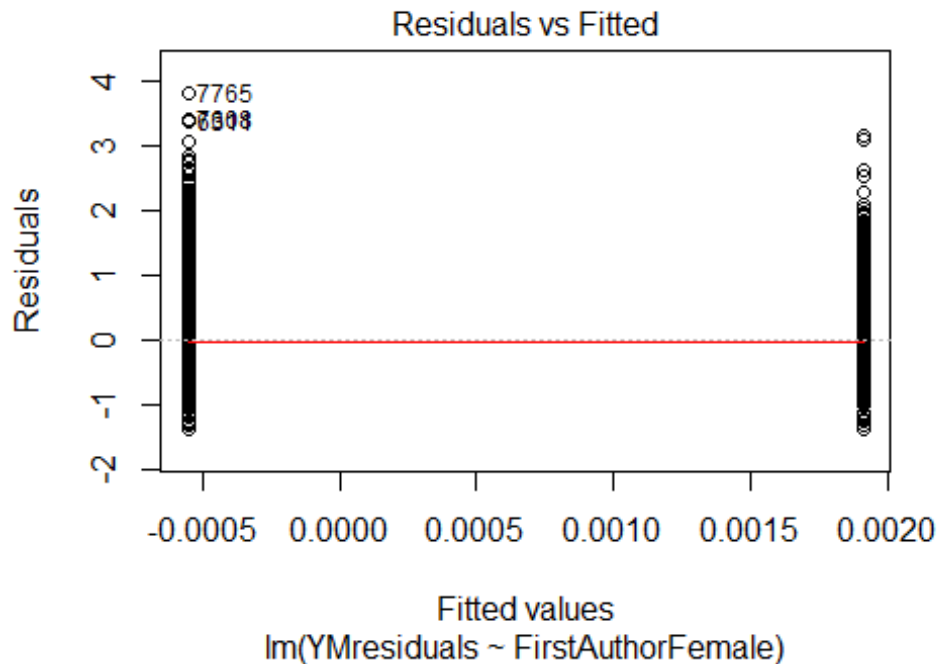
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 940"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1211"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 275 291 358 328 394 353 352 344 297 361 473 582 565 505 510
## 2011 2012
## 699 860
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 228 228 304 279 326 282 295 291 255 293 391 487 473 431 432
## 2011 2012
## 602 754
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 226 226 303 277 325 279 290 289 253 289 389 482 464 423 428
## 2011 2012
## 593 743
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 150, df = 16, p-value <2e-16

```



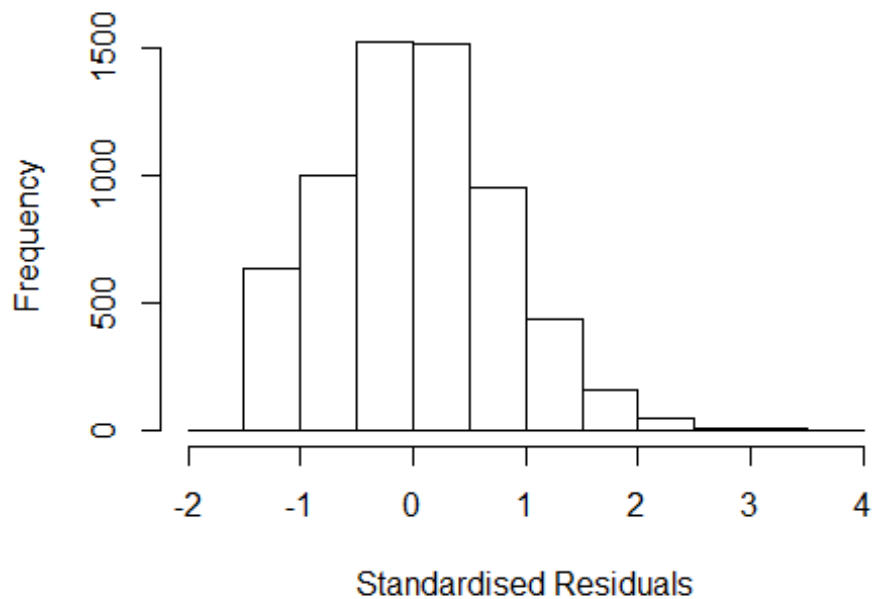
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0075, df = 1, p-value = 0.9
```





```
## [1] "Female first author team size 2018 geometric mean: 1.1580649269679"
## [1] "Male first author team size 2018 geometric mean: 1.07746552979718"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 80000, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.13422527400383"
## [1] "Male last author team size 2018 geometric mean: 1.08786634045307"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 77000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.392 1          2.096
## LastAuthorFemale  4.303 1          2.074
## UniqueAuthors    1.104 4          1.012
## Year              1.074 16          1.002
```

## Residuals from first and last author and team size



```
## [1] "List of 17 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 402    0040349812 3.709 1997    1211      1    2.789
## 1333  61249468539 3.822 2000    1208      3    2.875
## 2126  62449108692 3.929 2002    1211      1    2.589
## 2605  2142774713 4.094 2003    1211      1    2.878
## 2977  33846301138 4.379 2004    1211      1    3.127
## 6311  84855915732 4.594 2011    1208      4    3.443
## 6312  84855916060 4.298 2011    1208      4    3.161
## 6721  79953812056 3.867 2011    1211      1    2.716
## 7224  84880551540 3.867 2012    1211      1    2.625
## 7608  84872909005 4.721 2012    1202      4    3.266
## 7751  84455205704 3.943 2012    1211      1    2.686
## 7757  84855716574 3.757 2012    1211      1    2.500
## 7765  84869145733 5.134 2012    1211      1    3.877
## 7787  84855217343 3.954 2012    1207      2    2.712
## 7824  84856425812 3.956 2012    1211      3    2.699
## 7825  84856433228 4.077 2012    1211      3    2.629
## 7826  84856451242 4.492 2012    1211      3    3.250
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
```

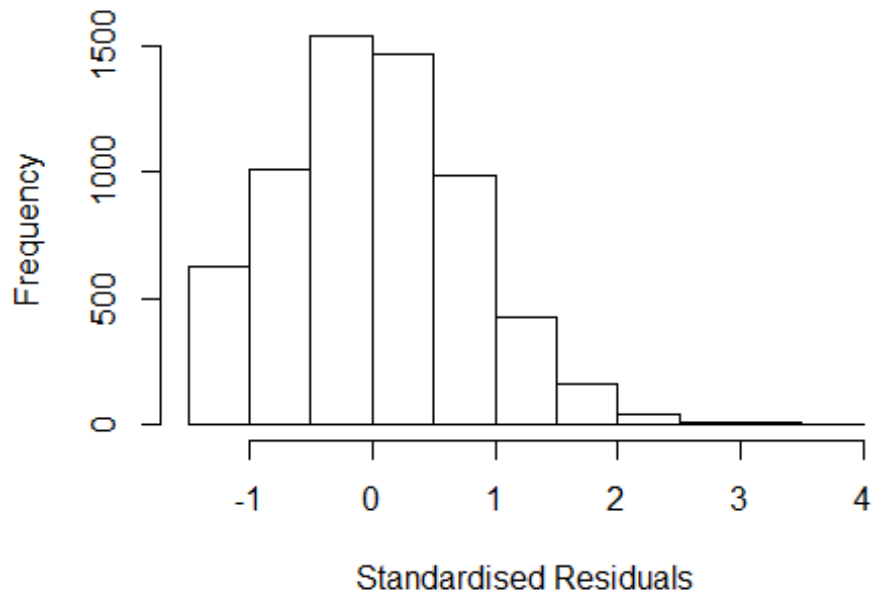
```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.53913 -0.52419 -0.00558  0.51868  3.87743
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.920325   0.049341  18.65 < 2e-16 ***
## FirstAuthorFemale1 -0.007546   0.050384  -0.15  0.88096
## LastAuthorFemale1 -0.006993   0.050684  -0.14  0.89026
## UniqueAuthors2    0.198870   0.040194   4.95  7.7e-07 ***
## UniqueAuthors3    0.253256   0.073240   3.46  0.00055 ***
## UniqueAuthors4    0.075623   0.142048   0.53  0.59448
## UniqueAuthors5    0.034913   0.135775   0.26  0.79708
## Year1997          -0.000743   0.068722  -0.01  0.99137
## Year1998           0.051445   0.063045   0.82  0.41452
## Year1999          -0.035432   0.062113  -0.57  0.56840
## Year2000           0.026995   0.061639   0.44  0.66143
## Year2001          -0.007120   0.066017  -0.11  0.91412
## Year2002           0.419934   0.074220   5.66  1.6e-08 ***
## Year2003           0.295978   0.070215   4.22  2.5e-05 ***
## Year2004           0.331394   0.071843   4.61  4.1e-06 ***
## Year2005           0.302891   0.070269   4.31  1.7e-05 ***
## Year2006           0.187548   0.063137   2.97  0.00298 **
## Year2007           0.187426   0.059118   3.17  0.00153 **
## Year2008           0.220312   0.061230   3.60  0.00032 ***
## Year2009           0.185549   0.060612   3.06  0.00221 **
## Year2010           0.171611   0.061654   2.78  0.00539 **
## Year2011           0.230863   0.060156   3.84  0.00013 ***
## Year2012           0.336247   0.060693   5.54  3.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.781
## Multiple R-squared:  0.0326, Adjusted R-squared:  0.0292
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 6197 is an outlier with |weight| = 0 ( < 1.6e-05);
## 501 weights are ~= 1. The remaining 5777 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0132 0.8700 0.9520 0.9120 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200

```

```
## trace.lev      mts compute.rd
##           0      1000         0
##           psi      subsampling      cov
##           "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.402 1      2.098
## LastAuthorFemale 4.399 1      2.097
## Year              1.022 16      1.001
```

### Residuals from first and last author



```
## [1] "List of 17 outliers with residuals above 2.5"
##           ScopusId NLCS Year OneField Fields residuals
## 402      0040349812 3.709 1997      1211      1      2.784
## 1333     61249468539 3.822 2000      1208      3      2.859
## 2126     62449108692 3.929 2002      1211      1      2.579
## 2605     2142774713 4.094 2003      1211      1      2.864
## 2977     33846301138 4.379 2004      1211      1      3.120
## 3039     33244477325 3.930 2004      1211      2      2.640
## 6311     84855915732 4.594 2011      1208      4      3.419
## 6312     84855916060 4.298 2011      1208      4      3.125
## 6721     79953812056 3.867 2011      1211      1      2.692
## 7224     84880551540 3.867 2012      1211      1      2.592
## 7608     84872909005 4.721 2012      1202      4      3.445
```

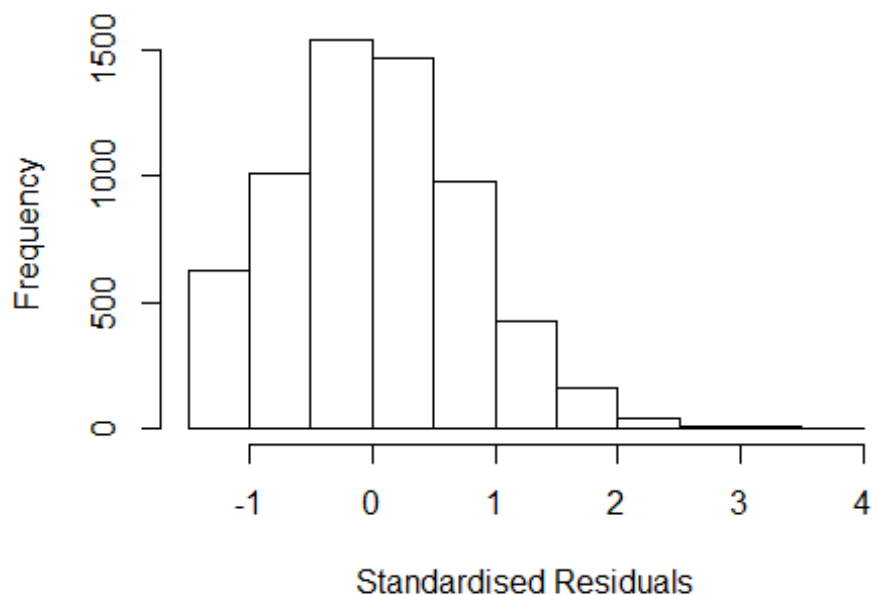
```

## 7751 84455205704 3.943 2012      1211      1      2.667
## 7765 84869145733 5.134 2012      1211      1      3.858
## 7787 84855217343 3.954 2012      1207      2      2.679
## 7824 84856425812 3.956 2012      1211      3      2.680
## 7825 84856433228 4.077 2012      1211      3      2.833
## 7826 84856451242 4.492 2012      1211      3      3.217
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.34991 -0.53867 -0.00836  0.53071  3.85764
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92784    0.04927   18.83 < 2e-16 ***
## FirstAuthorFemale1 0.03110    0.05072    0.61 0.53976
## LastAuthorFemale1 -0.03279    0.05155   -0.64 0.52478
## Year1997        -0.00321    0.06889   -0.05 0.96279
## Year1998         0.05053    0.06303    0.80 0.42275
## Year1999        -0.03314    0.06218   -0.53 0.59404
## Year2000         0.03467    0.06170    0.56 0.57426
## Year2001        -0.00273    0.06626   -0.04 0.96715
## Year2002         0.42206    0.07418    5.69 1.3e-08 ***
## Year2003         0.30167    0.07045    4.28 1.9e-05 ***
## Year2004         0.33100    0.07187    4.61 4.2e-06 ***
## Year2005         0.30754    0.07020    4.38 1.2e-05 ***
## Year2006         0.20015    0.06323    3.17 0.00156 **
## Year2007         0.19529    0.05905    3.31 0.00095 ***
## Year2008         0.22431    0.06118    3.67 0.00025 ***
## Year2009         0.20213    0.06056    3.34 0.00085 ***
## Year2010         0.18754    0.06162    3.04 0.00235 **
## Year2011         0.24719    0.05994    4.12 3.8e-05 ***
## Year2012         0.34852    0.06076    5.74 1.0e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.785
## Multiple R-squared:  0.027, Adjusted R-squared:  0.0242
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 6197 is an outlier with |weight| = 0 ( < 1.6e-05);
## 515 weights are ~ 1. The remaining 5763 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0152 0.8700 0.9520 0.9120 0.9860 0.9990
## Algorithmic parameters:

```

```
##          tuning.chi          bb          tuning.psi          refine.tol
##          1.55e+00          5.00e-01          4.69e+00          1.00e-07
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          1.59e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample          max.it          best.r.s          k.fast.s          k.max          maxit.scale
##          500          50          2          1          1000          200
## trace.lev          mts          compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.016 1          1.008
## Year          1.016 16          1.001
```

### Residuals from first author



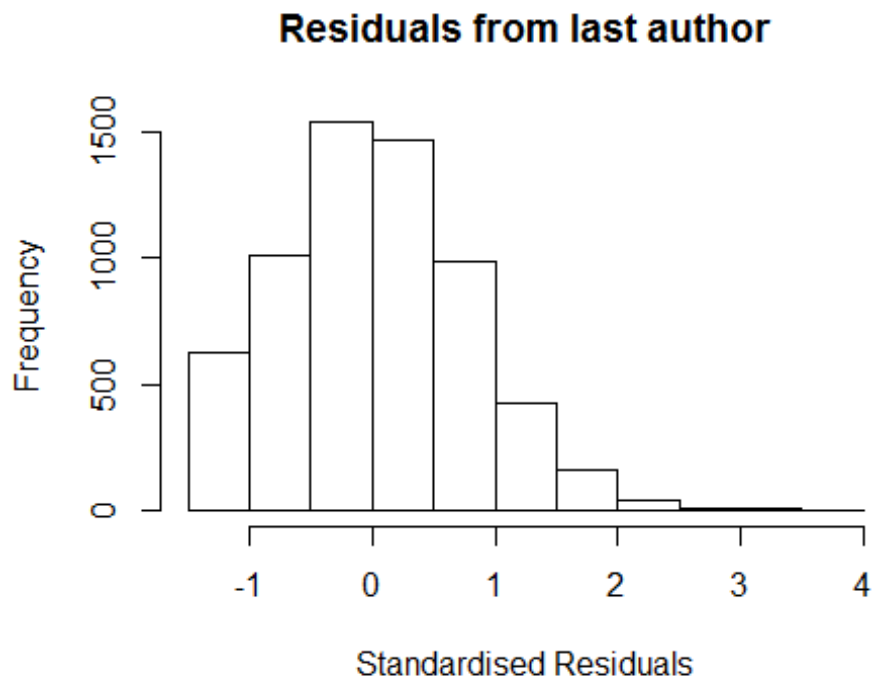
```
## [1] "List of 17 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 402 0040349812 3.709 1997 1211 1 2.784
## 1333 61249468539 3.822 2000 1208 3 2.859
## 2126 62449108692 3.929 2002 1211 1 2.579
## 2605 2142774713 4.094 2003 1211 1 2.864
## 2977 33846301138 4.379 2004 1211 1 3.120
```

```

## 3039 33244477325 3.930 2004      1211      2      2.640
## 6311 84855915732 4.594 2011      1208      4      3.419
## 6312 84855916060 4.298 2011      1208      4      3.125
## 6721 79953812056 3.867 2011      1211      1      2.692
## 7224 84880551540 3.867 2012      1211      1      2.592
## 7608 84872909005 4.721 2012      1202      4      3.445
## 7751 84455205704 3.943 2012      1211      1      2.667
## 7765 84869145733 5.134 2012      1211      1      3.858
## 7787 84855217343 3.954 2012      1207      2      2.679
## 7824 84856425812 3.956 2012      1211      3      2.680
## 7825 84856433228 4.077 2012      1211      3      2.833
## 7826 84856451242 4.492 2012      1211      3      3.217
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.35122 -0.53800 -0.00819  0.53055  3.85856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.92769    0.04927   18.83 < 2e-16 ***
## FirstAuthorFemale1 0.00181    0.02444    0.07  0.94083
## Year1997      -0.00365    0.06889   -0.05  0.95777
## Year1998       0.05007    0.06303    0.79  0.42697
## Year1999      -0.03394    0.06217   -0.55  0.58518
## Year2000       0.03469    0.06171    0.56  0.57407
## Year2001      -0.00282    0.06628   -0.04  0.96612
## Year2002       0.42171    0.07424    5.68 1.4e-08 ***
## Year2003       0.30137    0.07045    4.28 1.9e-05 ***
## Year2004       0.33039    0.07186    4.60 4.4e-06 ***
## Year2005       0.30666    0.07018    4.37 1.3e-05 ***
## Year2006       0.19940    0.06323    3.15 0.00162 **
## Year2007       0.19469    0.05905    3.30 0.00098 ***
## Year2008       0.22368    0.06118    3.66 0.00026 ***
## Year2009       0.20227    0.06058    3.34 0.00085 ***
## Year2010       0.18711    0.06161    3.04 0.00240 **
## Year2011       0.24701    0.05994    4.12 3.8e-05 ***
## Year2012       0.34775    0.06076    5.72 1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.785
## Multiple R-squared:  0.027, Adjusted R-squared:  0.0243
## Convergence in 10 IRWLS iterations
##
## Robustness weights:

```

```
## observation 6197 is an outlier with |weight| = 0 ( < 1.6e-05);
## 515 weights are ~= 1. The remaining 5763 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.015  0.871  0.952  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1      1.008
## Year      1.016 16      1.000
```



```
## [1] "List of 17 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
```



```

## 402 0040349812 3.709 1997 1211 1 2.784
## 1333 61249468539 3.822 2000 1208 3 2.859
## 2126 62449108692 3.929 2002 1211 1 2.579
## 2605 2142774713 4.094 2003 1211 1 2.864
## 2977 33846301138 4.379 2004 1211 1 3.120
## 3039 33244477325 3.930 2004 1211 2 2.640
## 6311 84855915732 4.594 2011 1208 4 3.419
## 6312 84855916060 4.298 2011 1208 4 3.125
## 6721 79953812056 3.867 2011 1211 1 2.692
## 7224 84880551540 3.867 2012 1211 1 2.592
## 7608 84872909005 4.721 2012 1202 4 3.445
## 7751 84455205704 3.943 2012 1211 1 2.667
## 7765 84869145733 5.134 2012 1211 1 3.858
## 7787 84855217343 3.954 2012 1207 2 2.679
## 7824 84856425812 3.956 2012 1211 3 2.680
## 7825 84856433228 4.077 2012 1211 3 2.833
## 7826 84856451242 4.492 2012 1211 3 3.217
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35069 -0.53847 -0.00749 0.53009 3.85681
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92860 0.04927 18.85 < 2e-16 ***
## LastAuthorFemale1 -0.00444 0.02478 -0.18 0.85793
## Year1997 -0.00345 0.06889 -0.05 0.96002
## Year1998 0.05018 0.06304 0.80 0.42601
## Year1999 -0.03352 0.06218 -0.54 0.58990
## Year2000 0.03490 0.06170 0.57 0.57171
## Year2001 -0.00237 0.06625 -0.04 0.97141
## Year2002 0.42210 0.07422 5.69 1.3e-08 ***
## Year2003 0.30160 0.07045 4.28 1.9e-05 ***
## Year2004 0.33064 0.07187 4.60 4.3e-06 ***
## Year2005 0.30698 0.07018 4.37 1.2e-05 ***
## Year2006 0.20007 0.06325 3.16 0.00157 **
## Year2007 0.19525 0.05905 3.31 0.00095 ***
## Year2008 0.22436 0.06119 3.67 0.00025 ***
## Year2009 0.20269 0.06059 3.35 0.00083 ***
## Year2010 0.18781 0.06162 3.05 0.00231 **
## Year2011 0.24743 0.05994 4.13 3.7e-05 ***
## Year2012 0.34860 0.06077 5.74 1.0e-08 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

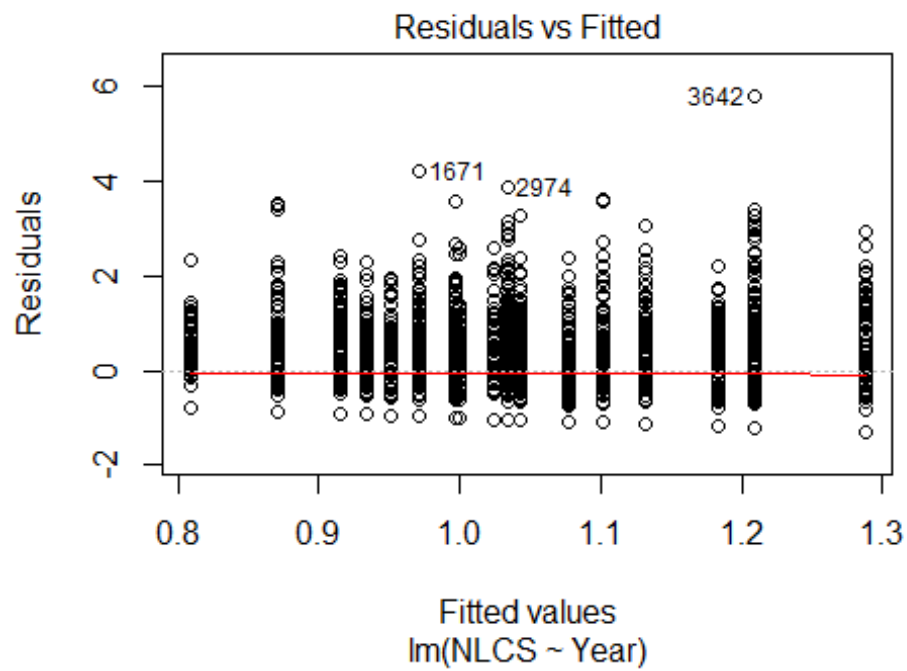
```

```

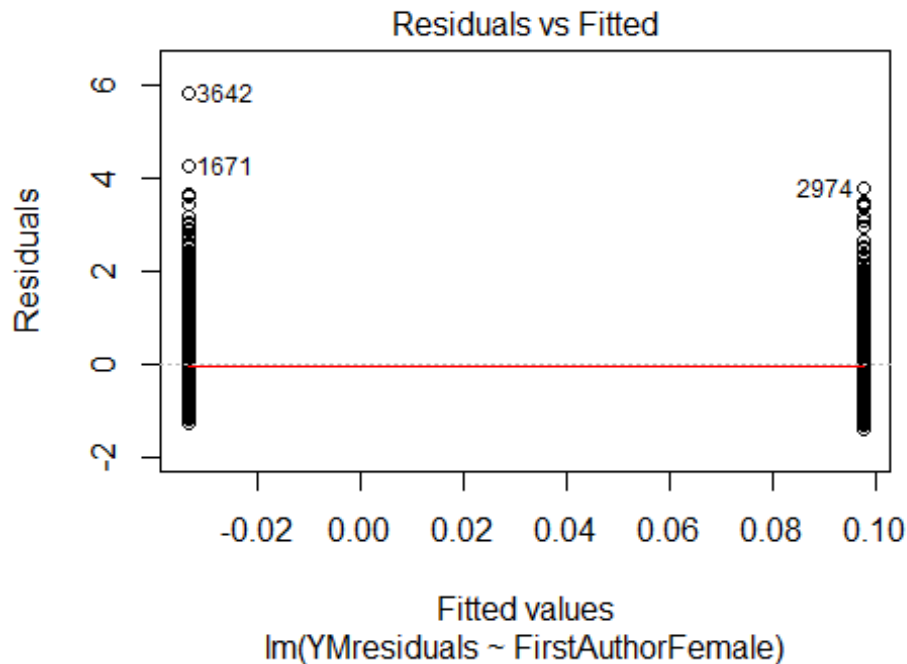
## Robust residual standard error: 0.785
## Multiple R-squared: 0.027, Adjusted R-squared: 0.0244
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 6197 is an outlier with |weight| = 0 ( < 1.6e-05);
## 511 weights are ~= 1. The remaining 5767 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0153 0.8710 0.9520 0.9120 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.59e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6279"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1212"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 146 149 160 181 162 194 150 192 190 262 238 268 239 257
## 2011 2012
## 403 395
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 105 112 130 131 153 131 168 129 171 173 238 200 233 213 223
## 2011 2012
## 354 326
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 104 110 129 131 151 131 167 129 169 173 237 198 231 211 221
## 2011 2012
## 346 323
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"

```

```
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 90, df = 16, p-value = 3e-12
```

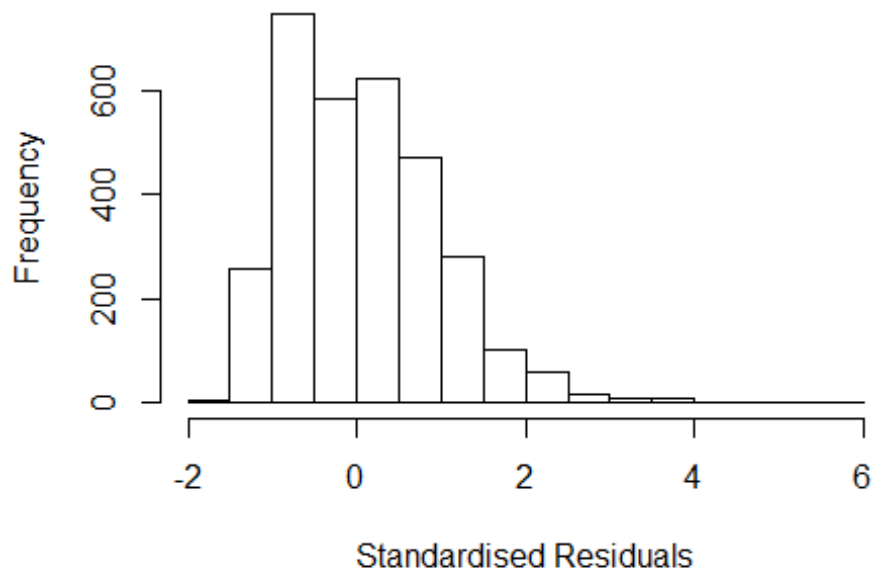


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.13042733790324"
## [1] "Male first author team size 2018 geometric mean: 1.09175010599033"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.14648285679905"
## [1] "Male last author team size 2018 geometric mean: 1.08546695225925"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 9.386 1          3.064
## LastAuthorFemale  9.215 1          3.036
## UniqueAuthors    1.246 4          1.028
## Year             1.114 16          1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 32 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 60    21344467895 3.616 1996    1212      1    2.597
## 356   85023947793 3.600 1998    1212      1    2.570
## 935   2442484904  4.711 2002    1212      1    3.728
## 1020  0036521981  3.836 2002    1212      1    2.757
## 1022  33646501384 4.667 2002    1212      1    3.588
## 1330  9144258932  4.244 2004    1202      2    3.037
## 1489  60950276825 4.177 2005    1212      1    3.041
## 1569  22144465521 3.686 2005    1212      1    2.646
## 1670  33751039356 3.732 2006    1212      1    2.876
## 1671  33751044705 5.186 2006    1212      1    3.901
## 1938  60950309686 4.316 2007    1212      1    2.823
## 2188  61149093449 4.559 2008    1212      1    3.669
## 2454  71649092061 4.388 2009    1212      1    3.559
## 2496  77950377327 4.274 2009    1212      2    3.540
## 2498  77950414533 4.382 2009    1212      2    3.553
## 2974  84860867265 4.913 2011    1212      1    3.486
## 3091  80052191693 3.870 2011    1212      1    2.568
## 3158  79959693303 3.799 2011    1212      1    2.926
## 3178  79957691598 3.938 2011    1212      1    3.065
## 3179  79958057053 4.177 2011    1212      1    3.304
## 3270  79952269370 4.063 2011    1212      1    3.190
## 3433  84868306956 3.686 2012    1208      2    2.650
## 3476  84866251389 3.965 2012    1212      1    2.929
## 3478  84863593550 4.631 2012    1212      1    3.595
## 3507  84927510827 4.286 2012    1212      1    3.155
```

```

## 3575 84861382392 4.380 2012      1212      1      2.799
## 3642 84857939525 7.007 2012      1212      1      5.971
## 3699 84866397895 3.965 2012      1212      1      2.929
## 3702 84871412120 4.080 2012      1212      1      3.044
## 3766 84856425812 3.956 2012      1211      3      2.920
## 3767 84856433228 4.077 2012      1211      3      2.632
## 3768 84856451242 4.492 2012      1211      3      3.361
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.65631 -0.81026 -0.00966  0.64862  5.97118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9234     0.0942   9.80  < 2e-16 ***
## FirstAuthorFemale1  0.1158     0.1131   1.02   0.306
## LastAuthorFemale1 -0.0203     0.1132  -0.18   0.858
## UniqueAuthors2     0.4292     0.0867   4.95 7.9e-07 ***
## UniqueAuthors3     0.4583     0.1787   2.57   0.010 *
## UniqueAuthors4     0.6933     0.3987   1.74   0.082 .
## UniqueAuthors5     0.4038     0.3887   1.04   0.299
## Year1997          -0.0526     0.1221  -0.43   0.667
## Year1998           0.0106     0.1155   0.09   0.927
## Year1999          -0.1896     0.1125  -1.69   0.092 .
## Year2000           0.0612     0.1115   0.55   0.583
## Year2001          -0.0613     0.1167  -0.53   0.600
## Year2002           0.0599     0.1184   0.51   0.613
## Year2003           0.2143     0.1257   1.71   0.088 .
## Year2004           0.2837     0.1241   2.29   0.022 *
## Year2005           0.1169     0.1209   0.97   0.334
## Year2006          -0.0678     0.1101  -0.62   0.538
## Year2007           0.0500     0.1108   0.45   0.652
## Year2008          -0.0334     0.1109  -0.30   0.763
## Year2009          -0.1898     0.1097  -1.73   0.084 .
## Year2010          -0.1132     0.1114  -1.02   0.310
## Year2011          -0.0501     0.1064  -0.47   0.638
## Year2012           0.1124     0.1111   1.01   0.312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.931
## Multiple R-squared:  0.0321, Adjusted R-squared:  0.0253
## Convergence in 12 IRWLS iterations

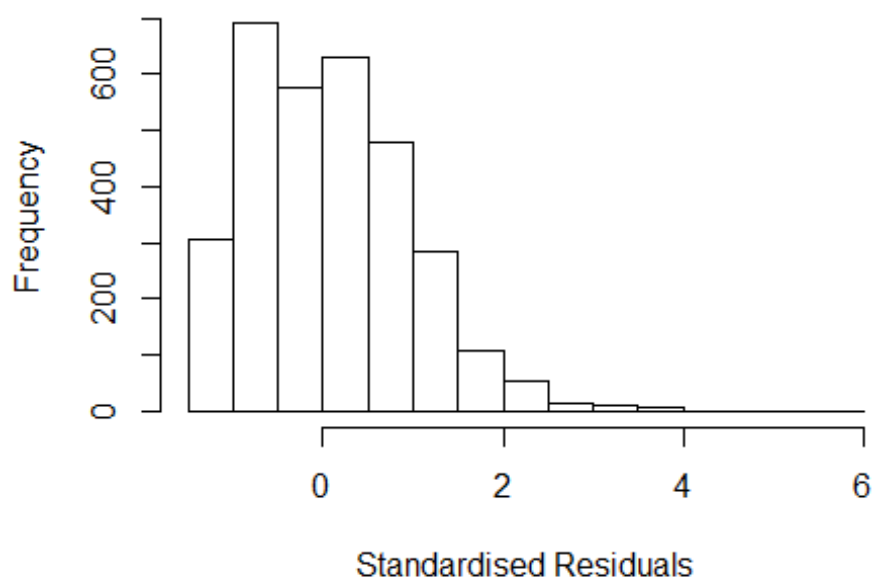
```

```

##
## Robustness weights:
## observation 3049 is an outlier with |weight| = 0 ( < 3.2e-05);
## 260 weights are ~= 1. The remaining 2900 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.040  0.896  0.932   0.915   0.985   0.999
## Algorithmic parameters:
##       tuning.chi                bb          tuning.psi          refine.tol
##       1.55e+00                5.00e-01          4.69e+00          1.00e-07
##       rel.tol                solve.tol          eps.outlier          eps.x
##       1.00e-07                1.00e-07          3.16e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##       5.00e-01                5.00e-01
##   nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##       500          50        2          1          1000        200
##   trace.lev      mts    compute.rd
##       0          1000        0
##           psi          subsampling          cov
##       "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##       "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 9.663 1          3.109
## LastAuthorFemale  9.668 1          3.109
## Year              1.037 16          1.001

```

## Residuals from first and last author



```
## [1] "List of 35 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 60      21344467895 3.616 1996      1212      1      2.563
## 356     85023947793 3.600 1998      1212      1      2.526
## 935     2442484904 4.711 2002      1212      1      3.716
## 1020    0036521981 3.836 2002      1212      1      2.718
## 1022    33646501384 4.667 2002      1212      1      3.549
## 1330    9144258932 4.244 2004      1202      2      3.031
## 1430    33244477325 3.930 2004      1211      2      2.522
## 1489    60950276825 4.177 2005      1212      1      3.008
## 1569    22144465521 3.686 2005      1212      1      2.640
## 1670    33751039356 3.732 2006      1212      1      2.858
## 1671    33751044705 5.186 2006      1212      1      4.312
## 1938    60950309686 4.316 2007      1212      1      3.142
## 2188    61149093449 4.559 2008      1212      1      3.651
## 2318    44349104204 3.664 2008      1212      1      2.756
## 2454    71649092061 4.388 2009      1212      1      3.523
## 2496    77950377327 4.274 2009      1212      2      3.532
## 2498    77950414533 4.382 2009      1212      2      3.517
## 2974    84860867265 4.913 2011      1212      1      3.880
## 3091    80052191693 3.870 2011      1212      1      2.960
## 3158    79959693303 3.799 2011      1212      1      2.889
## 3178    79957691598 3.938 2011      1212      1      3.028
## 3179    79958057053 4.177 2011      1212      1      3.267
## 3270    79952269370 4.063 2011      1212      1      3.153
## 3387    84872370373 3.707 2012      1212      1      2.736
## 3433    84868306956 3.686 2012      1208      2      2.642
```



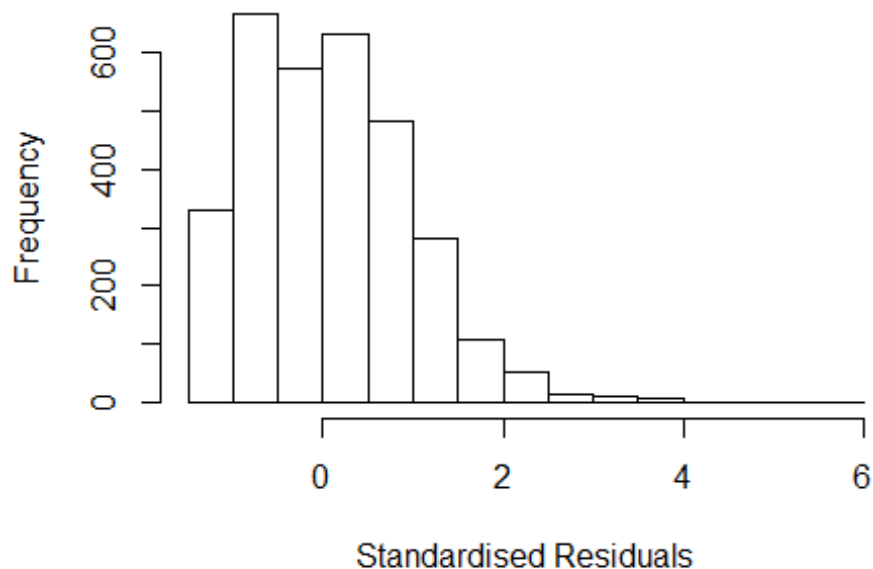
```

## 3476 84866251389 3.965 2012      1212      1      2.921
## 3478 84863593550 4.631 2012      1212      1      3.587
## 3507 84927510827 4.286 2012      1212      1      3.119
## 3575 84861382392 4.380 2012      1212      1      3.141
## 3642 84857939525 7.007 2012      1212      1      5.963
## 3699 84866397895 3.965 2012      1212      1      2.921
## 3702 84871412120 4.080 2012      1212      1      3.036
## 3766 84856425812 3.956 2012      1211      3      2.912
## 3767 84856433228 4.077 2012      1211      3      3.106
## 3768 84856451242 4.492 2012      1211      3      3.325
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q        Max
## -1.334973 -0.837255  0.000845  0.649929  5.962889
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9306      0.0937   9.93  <2e-16 ***
## FirstAuthorFemale1  0.1954      0.1160   1.68   0.092 .
## LastAuthorFemale1 -0.0729      0.1172  -0.62   0.534
## Year1997          -0.0434      0.1221  -0.36   0.722
## Year1998           0.0206      0.1149   0.18   0.858
## Year1999          -0.1766      0.1116  -1.58   0.114
## Year2000           0.0804      0.1109   0.73   0.468
## Year2001          -0.0530      0.1159  -0.46   0.647
## Year2002           0.0646      0.1174   0.55   0.582
## Year2003           0.2117      0.1249   1.70   0.090 .
## Year2004           0.2820      0.1241   2.27   0.023 *
## Year2005           0.1159      0.1204   0.96   0.336
## Year2006          -0.0563      0.1098  -0.51   0.608
## Year2007           0.0480      0.1102   0.44   0.663
## Year2008          -0.0225      0.1101  -0.20   0.838
## Year2009          -0.1881      0.1095  -1.72   0.086 .
## Year2010          -0.0933      0.1118  -0.83   0.404
## Year2011          -0.0204      0.1059  -0.19   0.847
## Year2012           0.1136      0.1113   1.02   0.308
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.937
## Multiple R-squared:  0.0199, Adjusted R-squared:  0.0143
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 3049 is an outlier with |weight| = 0 ( < 3.2e-05);

```

```
## 248 weights are ~= 1. The remaining 2912 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0012 0.8940 0.9350 0.9150 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.16e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001
```

### Residuals from first author



```
## [1] "List of 35 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 60      21344467895 3.616 1996      1212      1      2.563
```

```

## 356 85023947793 3.600 1998 1212 1 2.526
## 935 2442484904 4.711 2002 1212 1 3.716
## 1020 0036521981 3.836 2002 1212 1 2.718
## 1022 33646501384 4.667 2002 1212 1 3.549
## 1330 9144258932 4.244 2004 1202 2 3.031
## 1430 33244477325 3.930 2004 1211 2 2.522
## 1489 60950276825 4.177 2005 1212 1 3.008
## 1569 22144465521 3.686 2005 1212 1 2.640
## 1670 33751039356 3.732 2006 1212 1 2.858
## 1671 33751044705 5.186 2006 1212 1 4.312
## 1938 60950309686 4.316 2007 1212 1 3.142
## 2188 61149093449 4.559 2008 1212 1 3.651
## 2318 44349104204 3.664 2008 1212 1 2.756
## 2454 71649092061 4.388 2009 1212 1 3.523
## 2496 77950377327 4.274 2009 1212 2 3.532
## 2498 77950414533 4.382 2009 1212 2 3.517
## 2974 84860867265 4.913 2011 1212 1 3.880
## 3091 80052191693 3.870 2011 1212 1 2.960
## 3158 79959693303 3.799 2011 1212 1 2.889
## 3178 79957691598 3.938 2011 1212 1 3.028
## 3179 79958057053 4.177 2011 1212 1 3.267
## 3270 79952269370 4.063 2011 1212 1 3.153
## 3387 84872370373 3.707 2012 1212 1 2.736
## 3433 84868306956 3.686 2012 1208 2 2.642
## 3476 84866251389 3.965 2012 1212 1 2.921
## 3478 84863593550 4.631 2012 1212 1 3.587
## 3507 84927510827 4.286 2012 1212 1 3.119
## 3575 84861382392 4.380 2012 1212 1 3.141
## 3642 84857939525 7.007 2012 1212 1 5.963
## 3699 84866397895 3.965 2012 1212 1 2.921
## 3702 84871412120 4.080 2012 1212 1 3.036
## 3766 84856425812 3.956 2012 1211 3 2.912
## 3767 84856433228 4.077 2012 1211 3 3.106
## 3768 84856451242 4.492 2012 1211 3 3.325
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33788 -0.83652  0.00166  0.64505  5.96282
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9322     0.0937   9.95 < 2e-16 ***
## FirstAuthorFemale1 0.1270     0.0378   3.36 0.00079 ***
## Year1997        -0.0452     0.1221  -0.37 0.71098
## Year1998         0.0178     0.1148   0.16 0.87683

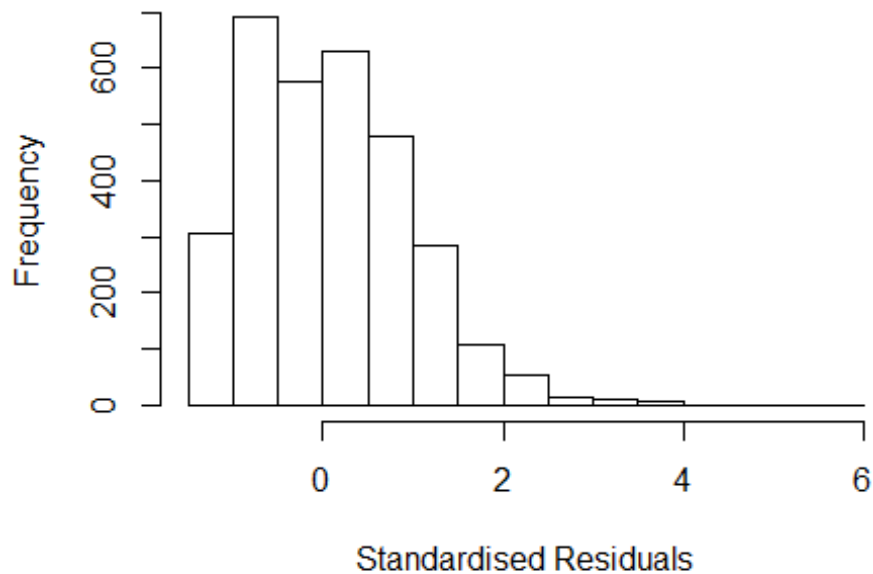
```

```

## Year1999          -0.1806      0.1115    -1.62  0.10551
## Year2000           0.0779      0.1109      0.70  0.48249
## Year2001          -0.0542      0.1158    -0.47  0.63976
## Year2002           0.0615      0.1173      0.52  0.60014
## Year2003           0.2096      0.1249      1.68  0.09322 .
## Year2004           0.2787      0.1238      2.25  0.02447 *
## Year2005           0.1136      0.1204      0.94  0.34544
## Year2006          -0.0584      0.1097    -0.53  0.59442
## Year2007           0.0447      0.1101      0.41  0.68450
## Year2008          -0.0247      0.1101    -0.22  0.82252
## Year2009          -0.1912      0.1094    -1.75  0.08070 .
## Year2010          -0.0957      0.1119    -0.86  0.39245
## Year2011          -0.0229      0.1059    -0.22  0.82912
## Year2012           0.1120      0.1114      1.01  0.31467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.938
## Multiple R-squared:  0.0197, Adjusted R-squared:  0.0144
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 3049 is an outlier with |weight| = 0 ( < 3.2e-05);
## 246 weights are ~= 1. The remaining 2914 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0014 0.8950 0.9350 0.9150 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.16e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1          1.012
## Year            1.024 16          1.001

```

## Residuals from last author



```
## [1] "List of 35 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 60    21344467895 3.616 1996    1212      1    2.563
## 356   85023947793 3.600 1998    1212      1    2.526
## 935   2442484904  4.711 2002    1212      1    3.716
## 1020  0036521981  3.836 2002    1212      1    2.718
## 1022  33646501384 4.667 2002    1212      1    3.549
## 1330  9144258932  4.244 2004    1202      2    3.031
## 1430  33244477325 3.930 2004    1211      2    2.522
## 1489  60950276825 4.177 2005    1212      1    3.008
## 1569  22144465521 3.686 2005    1212      1    2.640
## 1670  33751039356 3.732 2006    1212      1    2.858
## 1671  33751044705 5.186 2006    1212      1    4.312
## 1938  60950309686 4.316 2007    1212      1    3.142
## 2188  61149093449 4.559 2008    1212      1    3.651
## 2318  44349104204 3.664 2008    1212      1    2.756
## 2454  71649092061 4.388 2009    1212      1    3.523
## 2496  77950377327 4.274 2009    1212      2    3.532
## 2498  77950414533 4.382 2009    1212      2    3.517
## 2974  84860867265 4.913 2011    1212      1    3.880
## 3091  80052191693 3.870 2011    1212      1    2.960
## 3158  79959693303 3.799 2011    1212      1    2.889
## 3178  79957691598 3.938 2011    1212      1    3.028
## 3179  79958057053 4.177 2011    1212      1    3.267
## 3270  79952269370 4.063 2011    1212      1    3.153
## 3387  84872370373 3.707 2012    1212      1    2.736
## 3433  84868306956 3.686 2012    1208      2    2.642
```

```

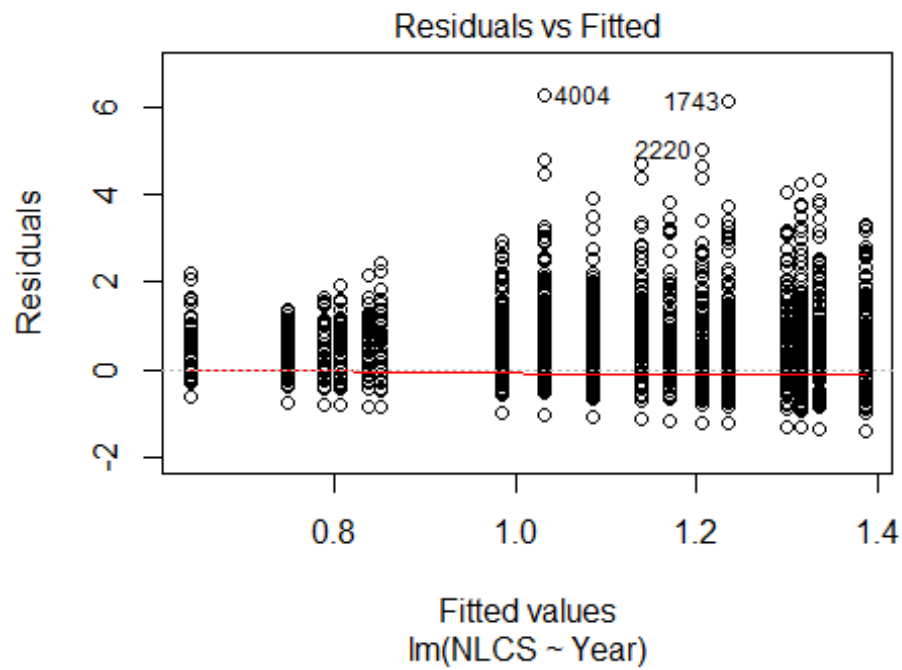
## 3476 84866251389 3.965 2012      1212      1      2.921
## 3478 84863593550 4.631 2012      1212      1      3.587
## 3507 84927510827 4.286 2012      1212      1      3.119
## 3575 84861382392 4.380 2012      1212      1      3.141
## 3642 84857939525 7.007 2012      1212      1      5.963
## 3699 84866397895 3.965 2012      1212      1      2.921
## 3702 84871412120 4.080 2012      1212      1      3.036
## 3766 84856425812 3.956 2012      1211      3      2.912
## 3767 84856433228 4.077 2012      1211      3      3.106
## 3768 84856451242 4.492 2012      1211      3      3.325
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.32500 -0.83999 -0.00219  0.64807  5.95707
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9412     0.0934   10.08  <2e-16 ***
## LastAuthorFemale1  0.1118     0.0381    2.94  0.0034 **
## Year1997         -0.0493     0.1220   -0.40  0.6861
## Year1998          0.0102     0.1145    0.09  0.9294
## Year1999         -0.1861     0.1115   -1.67  0.0954 .
## Year2000          0.0729     0.1108    0.66  0.5104
## Year2001         -0.0567     0.1157   -0.49  0.6238
## Year2002          0.0551     0.1173    0.47  0.6383
## Year2003          0.2058     0.1247    1.65  0.0989 .
## Year2004          0.2721     0.1236    2.20  0.0278 *
## Year2005          0.1095     0.1203    0.91  0.3628
## Year2006         -0.0621     0.1095   -0.57  0.5709
## Year2007          0.0390     0.1100    0.35  0.7226
## Year2008         -0.0293     0.1100   -0.27  0.7901
## Year2009         -0.1956     0.1092   -1.79  0.0733 .
## Year2010         -0.1012     0.1117   -0.91  0.3649
## Year2011         -0.0280     0.1057   -0.26  0.7914
## Year2012          0.1088     0.1112    0.98  0.3283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.94
## Multiple R-squared:  0.0188, Adjusted R-squared:  0.0135
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 3049 is an outlier with |weight| = 0 ( < 3.2e-05);
## 246 weights are ~ 1. The remaining 2914 ones are summarized as

```

```

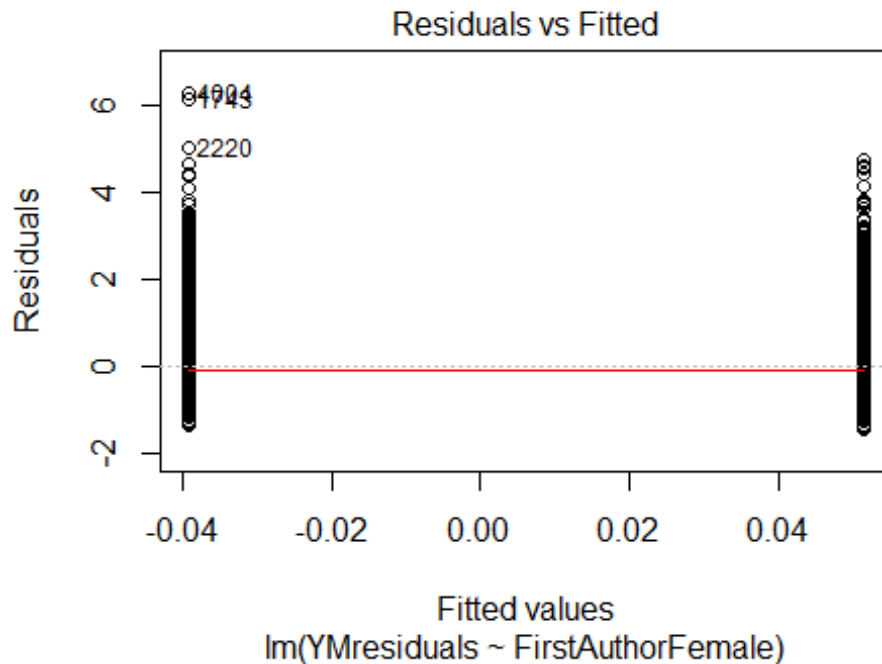
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0019 0.8950 0.9350 0.9160 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.16e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3161"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1213"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 111 130 115 99 108 104 240 198 243 320 314 335 351 239 330
## 2011 2012
## 431 610
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 93 96 89 83 94 91 213 178 216 286 272 307 309 215 286
## 2011 2012
## 378 536
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 92 95 89 82 93 90 211 176 212 283 267 296 305 210 283
## 2011 2012
## 370 525
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 170, df = 16, p-value <2e-16

```



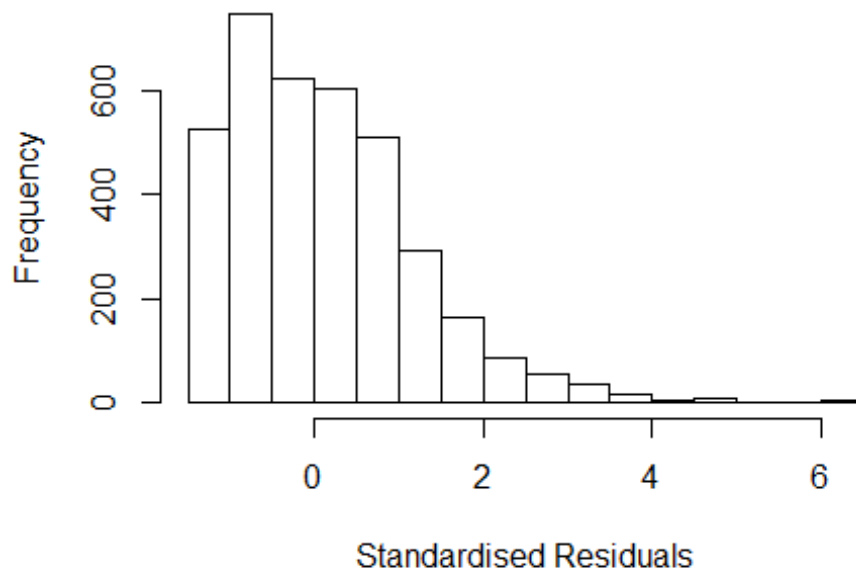
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.39, df = 1, p-value = 0.5
```





```
## [1] "Female first author team size 2018 geometric mean: 1.14064379483268"
## [1] "Male first author team size 2018 geometric mean: 1.11707913636303"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 39000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.17855865342353"
## [1] "Male last author team size 2018 geometric mean: 1.0713594916883"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 42000, p-value = 4e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.866 1          2.206
## LastAuthorFemale  4.865 1          2.206
## UniqueAuthors    1.074 4          1.009
## Year             1.098 16          1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 121 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 697 33750212545 3.982 2002    1213      1    2.958
## 698 34347213829 5.829 2002    1213      1    4.805
## 732 60950373677 3.920 2002    1208      2    3.003
## 800 1642297599 4.387 2002    1213      2    3.223
## 806 61449553775 3.982 2002    1213      1    3.065
## 808 85008525506 3.521 2002    1213      1    2.604
## 823 61949171102 3.665 2002    1208      2    2.748
## 845 84997942206 4.487 2002    1213      2    3.570
## 848 84998080695 3.794 2002    1213      2    2.877
## 849 84998183113 5.529 2002    1213      2    4.612
## 867 85007975083 3.980 2002    1213      2    3.063
## 942 21044457299 4.482 2003    1213      1    3.279
## 945 60950148528 4.356 2003    1213      1    3.260
## 959 27844509137 4.429 2003    1208      2    3.333
## 961 33645366634 4.044 2003    1208      2    2.841
## 1049 34248348215 5.357 2003    1208      2    4.261
## 1089 85007986198 3.649 2003    1213      2    2.553
## 1152 34249135977 5.191 2004    1213      1    3.970
## 1153 34249151392 3.874 2004    1213      1    2.653
## 1154 34249365472 4.381 2004    1213      1    3.160
## 1160 61449423891 3.666 2004    1213      1    2.551
## 1162 61449437192 5.096 2004    1213      1    3.716
## 1165 19644393451 3.827 2004    1208      2    2.712
## 1181 60950717875 3.827 2004    1208      2    2.712
## 1209 5044246868 3.743 2004    1213      3    2.522
```

## 1287	85012507421	3.874	2004	1213	1	2.653
## 1305	41749085870	5.674	2004	1208	2	4.559
## 1306	52949127270	3.675	2004	1208	2	2.560
## 1307	61149315889	4.521	2004	1208	2	3.300
## 1323	67649687426	3.675	2004	1208	2	2.560
## 1334	84998183199	3.693	2004	1213	2	2.578
## 1336	85008522337	4.779	2004	1213	2	3.558
## 1409	34249159351	4.052	2005	1213	1	2.827
## 1411	57349166039	4.052	2005	1213	1	2.827
## 1444	33745181909	4.105	2005	1210	2	2.880
## 1533	61449501605	4.172	2005	1213	2	2.789
## 1561	41549159477	3.933	2005	1213	2	2.708
## 1580	70350414060	4.001	2005	1208	2	2.776
## 1594	34249424493	4.685	2005	1213	1	3.460
## 1609	63849264187	4.576	2005	1213	1	3.245
## 1611	85008828873	4.199	2005	1213	1	2.868
## 1623	85012550597	4.199	2005	1213	1	2.974
## 1659	85008524788	3.997	2005	1213	2	2.595
## 1662	85008541082	4.668	2005	1213	2	3.196
## 1734	43249154616	3.958	2006	1213	1	2.915
## 1736	56749181819	4.476	2006	1213	1	3.433
## 1741	61149680861	4.620	2006	1213	1	3.577
## 1743	61449316310	7.375	2006	1213	1	6.332
## 1870	33748493765	4.219	2006	1210	2	3.176
## 1879	60949256671	3.958	2006	1213	1	2.915
## 1896	51249133709	4.147	2006	1213	1	3.104
## 1909	85008814567	4.147	2006	1213	1	2.998
## 1938	60949428823	4.945	2006	1208	2	3.796
## 1952	70449761232	4.357	2006	1208	2	3.208
## 1972	85008541069	4.250	2006	1213	2	3.207
## 2060	43249137313	4.625	2007	1213	1	3.563
## 2062	51249143307	3.622	2007	1213	1	2.666
## 2074	61049412681	3.622	2007	1213	1	2.666
## 2078	61449415197	4.625	2007	1213	1	3.563
## 2079	62749095981	4.089	2007	1213	1	3.027
## 2082	63849169071	3.622	2007	1213	1	2.560
## 2089	63849342094	4.089	2007	1213	1	3.133
## 2105	60949396166	3.834	2007	1208	2	2.878
## 2131	61949171840	3.629	2007	1208	2	2.673
## 2171	51249145194	3.460	2007	1202	6	2.504
## 2201	34250648363	3.919	2007	1213	2	2.857
## 2220	36048982830	6.200	2007	1210	2	5.244
## 2238	51249146700	3.870	2007	1213	1	2.808
## 2242	51249158469	4.089	2007	1213	1	3.133
## 2243	62449284443	4.625	2007	1213	1	3.669
## 2245	66249153136	5.835	2007	1213	1	4.702
## 2248	77954480003	5.575	2007	1213	1	4.619
## 2256	85012549979	5.835	2007	1213	1	4.773
## 2273	61149608982	3.629	2007	1208	2	2.567
## 2420	57749138210	3.939	2008	1213	1	2.791

##	2423	61049242834	4.598	2008	1213	1	3.450
##	2431	61449387618	5.541	2008	1213	1	4.287
##	2438	63849134726	5.079	2008	1213	1	3.931
##	2454	60950457792	4.816	2008	1213	2	3.668
##	2481	57749139264	3.783	2008	1213	3	2.529
##	2521	84973622246	4.091	2008	1206	3	2.757
##	2566	57049098170	5.016	2008	1210	2	3.709
##	2615	84904577020	4.298	2008	1213	1	3.150
##	2655	70449388215	4.476	2008	1213	2	3.063
##	2814	77950377327	4.274	2009	1212	2	3.263
##	2816	77950414533	4.382	2009	1212	2	3.264
##	2850	68349152348	4.361	2009	1213	1	3.350
##	2900	84863756017	3.856	2009	1213	1	2.738
##	2919	70449753866	4.612	2009	1213	2	3.601
##	2927	77949694173	4.988	2009	1208	2	3.870
##	3042	79955423767	3.903	2010	1210	2	2.621
##	3079	77955735971	3.545	2010	1213	3	2.690
##	3103	78249271418	3.777	2010	1210	2	2.922
##	3403	84857698666	4.563	2011	1208	2	3.514
##	3429	84855915732	4.594	2011	1208	4	3.651
##	3430	84855916060	4.298	2011	1208	4	3.249
##	3431	84855927603	3.619	2011	1208	4	2.676
##	3464	83755219444	4.311	2011	1213	1	3.262
##	3699	79953040564	3.817	2011	1213	1	2.768
##	3717	79956198385	4.972	2011	1208	2	3.923
##	3722	79952723945	3.604	2011	1213	3	2.555
##	3865	84871110200	3.607	2012	1210	2	2.777
##	3867	84871243321	4.302	2012	1208	2	3.472
##	3966	84870694677	4.195	2012	1213	1	3.365
##	3967	84870697316	3.615	2012	1213	1	2.785
##	3969	84870708659	3.615	2012	1213	1	2.679
##	3973	84872981155	3.615	2012	1213	1	2.679
##	4004	84866140479	7.287	2012	1208	2	6.457
##	4013	84868655179	3.559	2012	1208	2	2.729
##	4014	84868667202	3.962	2012	1208	2	3.026
##	4125	84861079400	3.962	2012	1208	2	3.026
##	4134	84865872698	5.502	2012	1213	2	4.566
##	4150	84868527748	4.195	2012	1213	1	3.365
##	4154	84868545918	4.195	2012	1213	1	3.259
##	4172	84859957258	3.520	2012	1202	4	2.690
##	4182	84861566134	3.608	2012	1213	3	2.672
##	4230	84863402857	3.926	2012	1213	1	2.990
##	4234	84856425818	5.835	2012	1208	2	4.899
##	4236	84856507096	3.559	2012	1208	2	2.729
##	4307	84869102856	3.615	2012	1213	1	2.679
##	4337	84865745109	3.962	2012	1208	2	3.026

##

## Call:

```
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
```

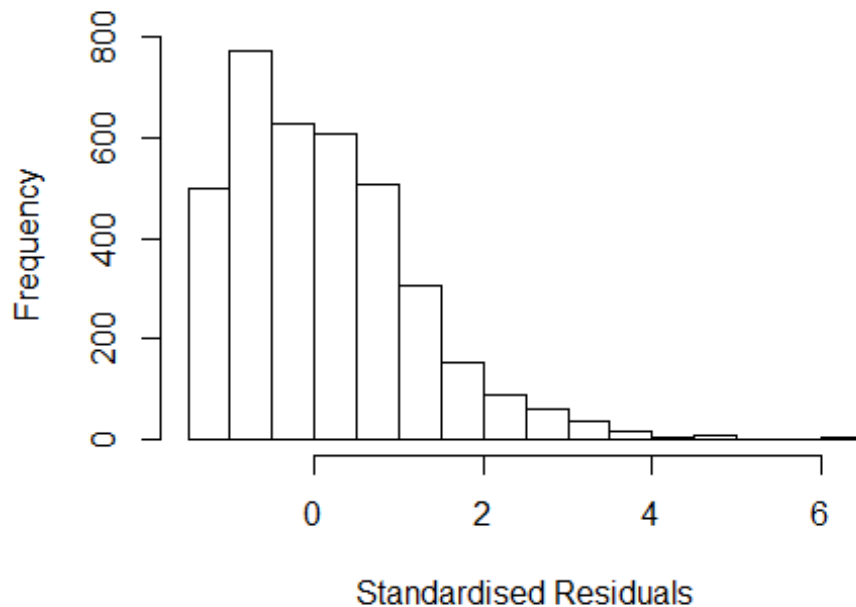
```

##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.4897 -0.8334 -0.0435  0.7250  6.4572
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7438     0.0986   7.54 5.8e-14 ***
## FirstAuthorFemale1 0.0880     0.0760   1.16 0.24645
## LastAuthorFemale1 0.0182     0.0760   0.24 0.81054
## UniqueAuthors2    0.1589     0.0641   2.48 0.01319 *
## UniqueAuthors3    0.1868     0.1196   1.56 0.11842
## UniqueAuthors4    0.4272     0.1518   2.81 0.00491 **
## UniqueAuthors5    0.0697     0.2182   0.32 0.74951
## Year1997          -0.1993     0.1223  -1.63 0.10332
## Year1998          -0.0602     0.1243  -0.48 0.62840
## Year1999          -0.0167     0.1366  -0.12 0.90293
## Year2000          -0.0117     0.1240  -0.09 0.92462
## Year2001          -0.0153     0.1279  -0.12 0.90476
## Year2002           0.1737     0.1252   1.39 0.16559
## Year2003           0.3526     0.1356   2.60 0.00936 **
## Year2004           0.3711     0.1302   2.85 0.00441 **
## Year2005           0.4808     0.1182   4.07 4.8e-05 ***
## Year2006           0.2989     0.1176   2.54 0.01108 *
## Year2007           0.2124     0.1165   1.82 0.06847 .
## Year2008           0.4039     0.1172   3.44 0.00058 ***
## Year2009           0.2675     0.1203   2.22 0.02626 *
## Year2010           0.1108     0.1100   1.01 0.31348
## Year2011           0.1987     0.1096   1.81 0.06995 .
## Year2012           0.0859     0.1074   0.80 0.42361
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 1
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.0258
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 7 observations c(543,1433,1834,1858,1868,3295,3479)
## are outliers with |weight| <= 5.8e-07 ( < 2.7e-05);
## 314 weights are ~= 1. The remaining 3358 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0013 0.8900 0.9360 0.9010 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x

```

```
##          1.00e-07          1.00e-07          2.72e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##          500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##          0          1000          0
##          psi      subsampling      cov
##          "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.639 1          2.154
## LastAuthorFemale 4.628 1          2.151
## Year              1.038 16          1.001
```

### Residuals from first and last author



```
## [1] "List of 123 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 697 33750212545 3.982 2002    1213      1      2.938
## 698 34347213829 5.829 2002    1213      1      4.785
## 732 60950373677 3.920 2002    1208      2      2.980
## 800 1642297599 4.387 2002    1213      2      3.359
## 806 61449553775 3.982 2002    1213      1      3.042
## 808 85008525506 3.521 2002    1213      1      2.581
```

## 823	61949171102	3.665	2002	1208	2	2.725
## 845	84997942206	4.487	2002	1213	2	3.547
## 847	84998021962	3.529	2002	1213	2	2.589
## 848	84998080695	3.794	2002	1213	2	2.854
## 849	84998183113	5.529	2002	1213	2	4.589
## 867	85007975083	3.980	2002	1213	2	3.040
## 942	21044457299	4.482	2003	1213	1	3.248
## 945	60950148528	4.356	2003	1213	1	3.226
## 959	27844509137	4.429	2003	1208	2	3.299
## 961	33645366634	4.044	2003	1208	2	2.810
## 1044	33749432390	3.807	2003	1208	2	2.589
## 1049	34248348215	5.357	2003	1208	2	4.227
## 1053	61049194206	3.761	2003	1202	2	2.543
## 1089	85007986198	3.649	2003	1213	2	2.519
## 1152	34249135977	5.191	2004	1213	1	3.961
## 1153	34249151392	3.874	2004	1213	1	2.644
## 1154	34249365472	4.381	2004	1213	1	3.151
## 1160	61449423891	3.666	2004	1213	1	2.540
## 1162	61449437192	5.096	2004	1213	1	3.866
## 1165	19644393451	3.827	2004	1208	2	2.701
## 1181	60950717875	3.827	2004	1208	2	2.701
## 1209	5044246868	3.743	2004	1213	3	2.513
## 1287	85012507421	3.874	2004	1213	1	2.644
## 1305	41749085870	5.674	2004	1208	2	4.548
## 1306	52949127270	3.675	2004	1208	2	2.549
## 1307	61149315889	4.521	2004	1208	2	3.291
## 1323	67649687426	3.675	2004	1208	2	2.549
## 1334	84998183199	3.693	2004	1213	2	2.567
## 1336	85008522337	4.779	2004	1213	2	3.549
## 1409	34249159351	4.052	2005	1213	1	2.803
## 1411	57349166039	4.052	2005	1213	1	2.803
## 1444	33745181909	4.105	2005	1210	2	2.856
## 1533	61449501605	4.172	2005	1213	2	2.923
## 1561	41549159477	3.933	2005	1213	2	2.684
## 1580	70350414060	4.001	2005	1208	2	2.752
## 1594	34249424493	4.685	2005	1213	1	3.436
## 1609	63849264187	4.576	2005	1213	1	3.223
## 1611	85008828873	4.199	2005	1213	1	2.846
## 1623	85012550597	4.199	2005	1213	1	2.950
## 1659	85008524788	3.997	2005	1213	2	2.731
## 1662	85008541082	4.668	2005	1213	2	3.331
## 1734	43249154616	3.958	2006	1213	1	2.897
## 1736	56749181819	4.476	2006	1213	1	3.415
## 1741	61149680861	4.620	2006	1213	1	3.559
## 1743	61449316310	7.375	2006	1213	1	6.314
## 1870	33748493765	4.219	2006	1210	2	3.158
## 1879	60949256671	3.958	2006	1213	1	2.897
## 1896	51249133709	4.147	2006	1213	1	3.086
## 1909	85008814567	4.147	2006	1213	1	2.982
## 1938	60949428823	4.945	2006	1208	2	3.780

##	1952	70449761232	4.357	2006	1208	2	3.192
##	1972	85008541069	4.250	2006	1213	2	3.189
##	2060	43249137313	4.625	2007	1213	1	3.543
##	2062	51249143307	3.622	2007	1213	1	2.645
##	2074	61049412681	3.622	2007	1213	1	2.645
##	2078	61449415197	4.625	2007	1213	1	3.543
##	2079	62749095981	4.089	2007	1213	1	3.007
##	2082	63849169071	3.622	2007	1213	1	2.540
##	2089	63849342094	4.089	2007	1213	1	3.112
##	2105	60949396166	3.834	2007	1208	2	2.857
##	2131	61949171840	3.629	2007	1208	2	2.652
##	2201	34250648363	3.919	2007	1213	2	2.837
##	2220	36048982830	6.200	2007	1210	2	5.223
##	2238	51249146700	3.870	2007	1213	1	2.788
##	2242	51249158469	4.089	2007	1213	1	3.112
##	2243	62449284443	4.625	2007	1213	1	3.648
##	2245	66249153136	5.835	2007	1213	1	4.841
##	2248	77954480003	5.575	2007	1213	1	4.598
##	2256	85012549979	5.835	2007	1213	1	4.753
##	2273	61149608982	3.629	2007	1208	2	2.547
##	2420	57749138210	3.939	2008	1213	1	2.771
##	2423	61049242834	4.598	2008	1213	1	3.430
##	2431	61449387618	5.541	2008	1213	1	4.269
##	2438	63849134726	5.079	2008	1213	1	3.911
##	2454	60950457792	4.816	2008	1213	2	3.648
##	2481	57749139264	3.783	2008	1213	3	2.511
##	2521	84973622246	4.091	2008	1206	3	2.923
##	2566	57049098170	5.016	2008	1210	2	3.848
##	2615	84904577020	4.298	2008	1213	1	3.130
##	2655	70449388215	4.476	2008	1213	2	3.204
##	2814	77950377327	4.274	2009	1212	2	3.246
##	2816	77950414533	4.382	2009	1212	2	3.250
##	2850	68349152348	4.361	2009	1213	1	3.333
##	2900	84863756017	3.856	2009	1213	1	2.724
##	2919	70449753866	4.612	2009	1213	2	3.584
##	2927	77949694173	4.988	2009	1208	2	3.856
##	3042	79955423767	3.903	2010	1210	2	3.013
##	3079	77955735971	3.545	2010	1213	3	2.655
##	3103	78249271418	3.777	2010	1210	2	2.887
##	3403	84857698666	4.563	2011	1208	2	3.495
##	3429	84855915732	4.594	2011	1208	4	3.630
##	3430	84855916060	4.298	2011	1208	4	3.230
##	3431	84855927603	3.619	2011	1208	4	2.655
##	3464	83755219444	4.311	2011	1213	1	3.243
##	3699	79953040564	3.817	2011	1213	1	2.749
##	3717	79956198385	4.972	2011	1208	2	3.904
##	3722	79952723945	3.604	2011	1213	3	2.536
##	3865	84871110200	3.607	2012	1210	2	2.759
##	3867	84871243321	4.302	2012	1208	2	3.454
##	3966	84870694677	4.195	2012	1213	1	3.347



```

## 3967 84870697316 3.615 2012      1213      1      2.767
## 3969 84870708659 3.615 2012      1213      1      2.663
## 3973 84872981155 3.615 2012      1213      1      2.663
## 4004 84866140479 7.287 2012      1208      2      6.439
## 4013 84868655179 3.559 2012      1208      2      2.711
## 4014 84868667202 3.962 2012      1208      2      3.010
## 4125 84861079400 3.962 2012      1208      2      3.010
## 4134 84865872698 5.502 2012      1213      2      4.550
## 4150 84868527748 4.195 2012      1213      1      3.347
## 4154 84868545918 4.195 2012      1213      1      3.243
## 4172 84859957258 3.520 2012      1202      4      2.672
## 4182 84861566134 3.608 2012      1213      3      2.656
## 4230 84863402857 3.926 2012      1213      1      2.974
## 4234 84856425818 5.835 2012      1208      2      4.883
## 4236 84856507096 3.559 2012      1208      2      2.711
## 4307 84869102856 3.615 2012      1213      1      2.663
## 4337 84865745109 3.962 2012      1208      2      3.010
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3533 -0.8480 -0.0491  0.7213  6.4390
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7652     0.0987   7.75 1.2e-14 ***
## FirstAuthorFemale1  0.0875     0.0744   1.18 0.23937
## LastAuthorFemale1  0.0165     0.0743   0.22 0.82383
## Year1997          -0.1984     0.1225  -1.62 0.10536
## Year1998          -0.0689     0.1240  -0.56 0.57853
## Year1999          -0.0192     0.1371  -0.14 0.88857
## Year2000          -0.0155     0.1244  -0.12 0.90106
## Year2001          -0.0162     0.1291  -0.13 0.90015
## Year2002           0.1752     0.1264   1.39 0.16568
## Year2003           0.3648     0.1360   2.68 0.00735 **
## Year2004           0.3608     0.1300   2.78 0.00553 **
## Year2005           0.4841     0.1184   4.09 4.4e-05 ***
## Year2006           0.2959     0.1180   2.51 0.01218 *
## Year2007           0.2123     0.1170   1.81 0.06968 .
## Year2008           0.4031     0.1175   3.43 0.00061 ***
## Year2009           0.2629     0.1208   2.18 0.02956 *
## Year2010           0.1248     0.1106   1.13 0.25904
## Year2011           0.1991     0.1102   1.81 0.07102 .
## Year2012           0.0828     0.1080   0.77 0.44287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

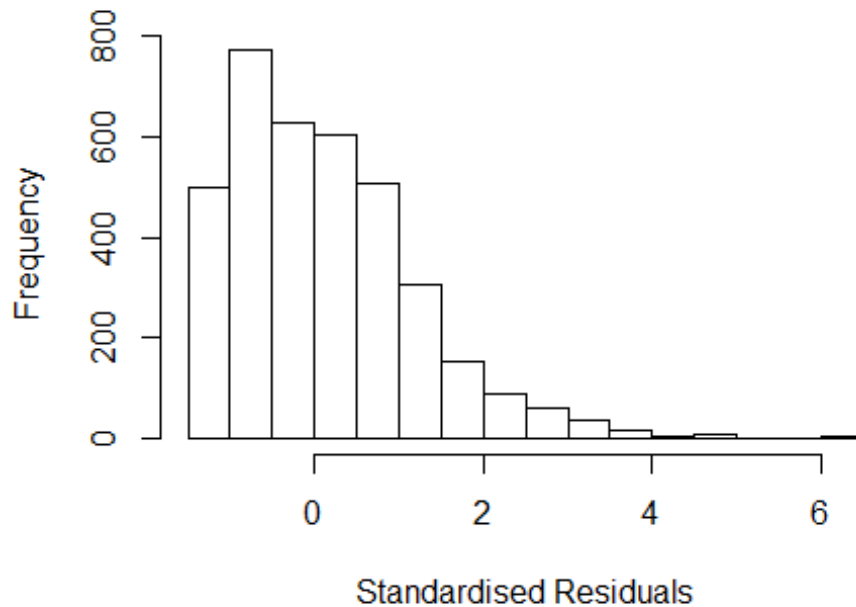
```

```

##
## Robust residual standard error: 1.01
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.0233
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 7 observations c(543,1433,1834,1858,1868,3295,3479)
## are outliers with |weight| = 0 ( < 2.7e-05);
## 312 weights are ~= 1. The remaining 3360 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0032 0.8890 0.9370 0.9020 0.9810 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.72e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.019 1           1.009
## Year              1.019 16           1.001

```

## Residuals from first author



```
## [1] "List of 123 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 697  33750212545 3.982 2002    1213     1    2.938
## 698  34347213829 5.829 2002    1213     1    4.785
## 732  60950373677 3.920 2002    1208     2    2.980
## 800   1642297599 4.387 2002    1213     2    3.359
## 806  61449553775 3.982 2002    1213     1    3.042
## 808  85008525506 3.521 2002    1213     1    2.581
## 823  61949171102 3.665 2002    1208     2    2.725
## 845  84997942206 4.487 2002    1213     2    3.547
## 847  84998021962 3.529 2002    1213     2    2.589
## 848  84998080695 3.794 2002    1213     2    2.854
## 849  84998183113 5.529 2002    1213     2    4.589
## 867  85007975083 3.980 2002    1213     2    3.040
## 942  21044457299 4.482 2003    1213     1    3.248
## 945  60950148528 4.356 2003    1213     1    3.226
## 959  27844509137 4.429 2003    1208     2    3.299
## 961  33645366634 4.044 2003    1208     2    2.810
## 1044 33749432390 3.807 2003    1208     2    2.589
## 1049 34248348215 5.357 2003    1208     2    4.227
## 1053 61049194206 3.761 2003    1202     2    2.543
## 1089 85007986198 3.649 2003    1213     2    2.519
## 1152 34249135977 5.191 2004    1213     1    3.961
## 1153 34249151392 3.874 2004    1213     1    2.644
## 1154 34249365472 4.381 2004    1213     1    3.151
## 1160 61449423891 3.666 2004    1213     1    2.540
## 1162 61449437192 5.096 2004    1213     1    3.866
```

##	1165	19644393451	3.827	2004	1208	2	2.701
##	1181	60950717875	3.827	2004	1208	2	2.701
##	1209	5044246868	3.743	2004	1213	3	2.513
##	1287	85012507421	3.874	2004	1213	1	2.644
##	1305	41749085870	5.674	2004	1208	2	4.548
##	1306	52949127270	3.675	2004	1208	2	2.549
##	1307	61149315889	4.521	2004	1208	2	3.291
##	1323	67649687426	3.675	2004	1208	2	2.549
##	1334	84998183199	3.693	2004	1213	2	2.567
##	1336	85008522337	4.779	2004	1213	2	3.549
##	1409	34249159351	4.052	2005	1213	1	2.803
##	1411	57349166039	4.052	2005	1213	1	2.803
##	1444	33745181909	4.105	2005	1210	2	2.856
##	1533	61449501605	4.172	2005	1213	2	2.923
##	1561	41549159477	3.933	2005	1213	2	2.684
##	1580	70350414060	4.001	2005	1208	2	2.752
##	1594	34249424493	4.685	2005	1213	1	3.436
##	1609	63849264187	4.576	2005	1213	1	3.223
##	1611	85008828873	4.199	2005	1213	1	2.846
##	1623	85012550597	4.199	2005	1213	1	2.950
##	1659	85008524788	3.997	2005	1213	2	2.731
##	1662	85008541082	4.668	2005	1213	2	3.331
##	1734	43249154616	3.958	2006	1213	1	2.897
##	1736	56749181819	4.476	2006	1213	1	3.415
##	1741	61149680861	4.620	2006	1213	1	3.559
##	1743	61449316310	7.375	2006	1213	1	6.314
##	1870	33748493765	4.219	2006	1210	2	3.158
##	1879	60949256671	3.958	2006	1213	1	2.897
##	1896	51249133709	4.147	2006	1213	1	3.086
##	1909	85008814567	4.147	2006	1213	1	2.982
##	1938	60949428823	4.945	2006	1208	2	3.780
##	1952	70449761232	4.357	2006	1208	2	3.192
##	1972	85008541069	4.250	2006	1213	2	3.189
##	2060	43249137313	4.625	2007	1213	1	3.543
##	2062	51249143307	3.622	2007	1213	1	2.645
##	2074	61049412681	3.622	2007	1213	1	2.645
##	2078	61449415197	4.625	2007	1213	1	3.543
##	2079	62749095981	4.089	2007	1213	1	3.007
##	2082	63849169071	3.622	2007	1213	1	2.540
##	2089	63849342094	4.089	2007	1213	1	3.112
##	2105	60949396166	3.834	2007	1208	2	2.857
##	2131	61949171840	3.629	2007	1208	2	2.652
##	2201	34250648363	3.919	2007	1213	2	2.837
##	2220	36048982830	6.200	2007	1210	2	5.223
##	2238	51249146700	3.870	2007	1213	1	2.788
##	2242	51249158469	4.089	2007	1213	1	3.112
##	2243	62449284443	4.625	2007	1213	1	3.648
##	2245	66249153136	5.835	2007	1213	1	4.841
##	2248	77954480003	5.575	2007	1213	1	4.598
##	2256	85012549979	5.835	2007	1213	1	4.753

##	2273	61149608982	3.629	2007	1208	2	2.547
##	2420	57749138210	3.939	2008	1213	1	2.771
##	2423	61049242834	4.598	2008	1213	1	3.430
##	2431	61449387618	5.541	2008	1213	1	4.269
##	2438	63849134726	5.079	2008	1213	1	3.911
##	2454	60950457792	4.816	2008	1213	2	3.648
##	2481	57749139264	3.783	2008	1213	3	2.511
##	2521	84973622246	4.091	2008	1206	3	2.923
##	2566	57049098170	5.016	2008	1210	2	3.848
##	2615	84904577020	4.298	2008	1213	1	3.130
##	2655	70449388215	4.476	2008	1213	2	3.204
##	2814	77950377327	4.274	2009	1212	2	3.246
##	2816	77950414533	4.382	2009	1212	2	3.250
##	2850	68349152348	4.361	2009	1213	1	3.333
##	2900	84863756017	3.856	2009	1213	1	2.724
##	2919	70449753866	4.612	2009	1213	2	3.584
##	2927	77949694173	4.988	2009	1208	2	3.856
##	3042	79955423767	3.903	2010	1210	2	3.013
##	3079	77955735971	3.545	2010	1213	3	2.655
##	3103	78249271418	3.777	2010	1210	2	2.887
##	3403	84857698666	4.563	2011	1208	2	3.495
##	3429	84855915732	4.594	2011	1208	4	3.630
##	3430	84855916060	4.298	2011	1208	4	3.230
##	3431	84855927603	3.619	2011	1208	4	2.655
##	3464	83755219444	4.311	2011	1213	1	3.243
##	3699	79953040564	3.817	2011	1213	1	2.749
##	3717	79956198385	4.972	2011	1208	2	3.904
##	3722	79952723945	3.604	2011	1213	3	2.536
##	3865	84871110200	3.607	2012	1210	2	2.759
##	3867	84871243321	4.302	2012	1208	2	3.454
##	3966	84870694677	4.195	2012	1213	1	3.347
##	3967	84870697316	3.615	2012	1213	1	2.767
##	3969	84870708659	3.615	2012	1213	1	2.663
##	3973	84872981155	3.615	2012	1213	1	2.663
##	4004	84866140479	7.287	2012	1208	2	6.439
##	4013	84868655179	3.559	2012	1208	2	2.711
##	4014	84868667202	3.962	2012	1208	2	3.010
##	4125	84861079400	3.962	2012	1208	2	3.010
##	4134	84865872698	5.502	2012	1213	2	4.550
##	4150	84868527748	4.195	2012	1213	1	3.347
##	4154	84868545918	4.195	2012	1213	1	3.243
##	4172	84859957258	3.520	2012	1202	4	2.672
##	4182	84861566134	3.608	2012	1213	3	2.656
##	4230	84863402857	3.926	2012	1213	1	2.974
##	4234	84856425818	5.835	2012	1208	2	4.883
##	4236	84856507096	3.559	2012	1208	2	2.711
##	4307	84869102856	3.615	2012	1213	1	2.663
##	4337	84865745109	3.962	2012	1208	2	3.010

##

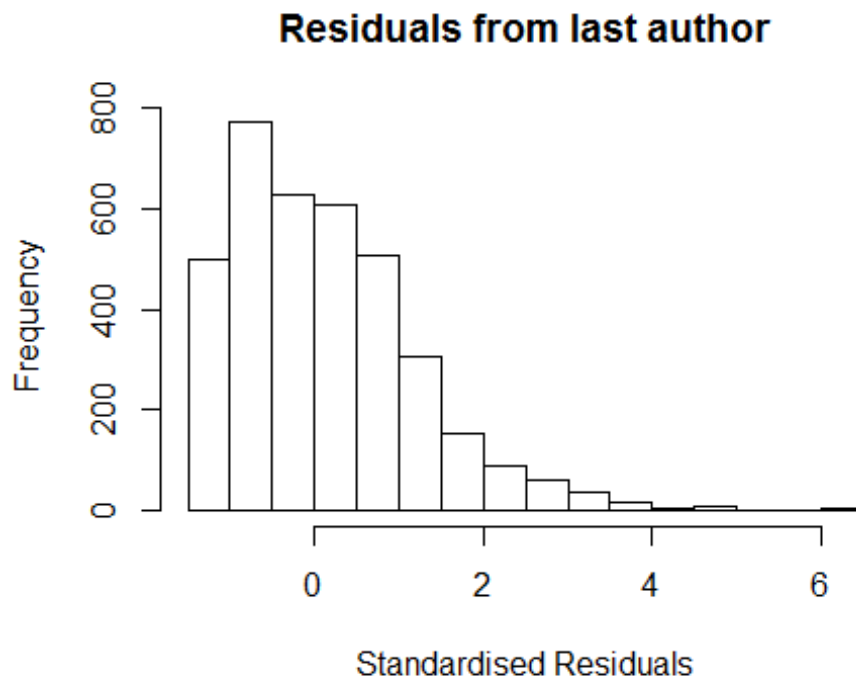
## Call:

```

## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.353 -0.848 -0.046  0.719  6.439
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7658     0.0986   7.76 1.1e-14 ***
## FirstAuthorFemale1 0.1025     0.0349   2.94 0.00333 **
## Year1997         -0.1986     0.1225  -1.62 0.10508
## Year1998         -0.0688     0.1240  -0.56 0.57890
## Year1999         -0.0188     0.1371  -0.14 0.89087
## Year2000         -0.0162     0.1243  -0.13 0.89603
## Year2001         -0.0164     0.1291  -0.13 0.89910
## Year2002          0.1753     0.1264   1.39 0.16553
## Year2003          0.3642     0.1360   2.68 0.00745 **
## Year2004          0.3611     0.1300   2.78 0.00549 **
## Year2005          0.4842     0.1184   4.09 4.4e-05 ***
## Year2006          0.2960     0.1180   2.51 0.01218 *
## Year2007          0.2122     0.1170   1.81 0.06976 .
## Year2008          0.4031     0.1176   3.43 0.00061 ***
## Year2009          0.2633     0.1207   2.18 0.02925 *
## Year2010          0.1252     0.1105   1.13 0.25738
## Year2011          0.1990     0.1102   1.81 0.07110 .
## Year2012          0.0827     0.1080   0.77 0.44386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 1.01
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.0236
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 7 observations c(543,1433,1834,1858,1868,3295,3479)
## are outliers with |weight| = 0 ( < 2.7e-05);
## 313 weights are ~= 1. The remaining 3359 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0032 0.8890 0.9370 0.9020 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200

```

```
##      trace.lev      mts  compute.rd
##              0      1000          0
##              psi      subsampling      cov
##              "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##              "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1      1.008
## Year              1.017 16      1.001
```



```
## [1] "List of 123 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 697  33750212545 3.982 2002    1213      1    2.938
## 698  34347213829 5.829 2002    1213      1    4.785
## 732  60950373677 3.920 2002    1208      2    2.980
## 800   1642297599 4.387 2002    1213      2    3.359
## 806  61449553775 3.982 2002    1213      1    3.042
## 808  85008525506 3.521 2002    1213      1    2.581
## 823  61949171102 3.665 2002    1208      2    2.725
## 845  84997942206 4.487 2002    1213      2    3.547
## 847  84998021962 3.529 2002    1213      2    2.589
## 848  84998080695 3.794 2002    1213      2    2.854
## 849  84998183113 5.529 2002    1213      2    4.589
## 867  85007975083 3.980 2002    1213      2    3.040
## 942  21044457299 4.482 2003    1213      1    3.248
```

## 945	60950148528	4.356	2003	1213	1	3.226
## 959	27844509137	4.429	2003	1208	2	3.299
## 961	33645366634	4.044	2003	1208	2	2.810
## 1044	33749432390	3.807	2003	1208	2	2.589
## 1049	34248348215	5.357	2003	1208	2	4.227
## 1053	61049194206	3.761	2003	1202	2	2.543
## 1089	85007986198	3.649	2003	1213	2	2.519
## 1152	34249135977	5.191	2004	1213	1	3.961
## 1153	34249151392	3.874	2004	1213	1	2.644
## 1154	34249365472	4.381	2004	1213	1	3.151
## 1160	61449423891	3.666	2004	1213	1	2.540
## 1162	61449437192	5.096	2004	1213	1	3.866
## 1165	19644393451	3.827	2004	1208	2	2.701
## 1181	60950717875	3.827	2004	1208	2	2.701
## 1209	5044246868	3.743	2004	1213	3	2.513
## 1287	85012507421	3.874	2004	1213	1	2.644
## 1305	41749085870	5.674	2004	1208	2	4.548
## 1306	52949127270	3.675	2004	1208	2	2.549
## 1307	61149315889	4.521	2004	1208	2	3.291
## 1323	67649687426	3.675	2004	1208	2	2.549
## 1334	84998183199	3.693	2004	1213	2	2.567
## 1336	85008522337	4.779	2004	1213	2	3.549
## 1409	34249159351	4.052	2005	1213	1	2.803
## 1411	57349166039	4.052	2005	1213	1	2.803
## 1444	33745181909	4.105	2005	1210	2	2.856
## 1533	61449501605	4.172	2005	1213	2	2.923
## 1561	41549159477	3.933	2005	1213	2	2.684
## 1580	70350414060	4.001	2005	1208	2	2.752
## 1594	34249424493	4.685	2005	1213	1	3.436
## 1609	63849264187	4.576	2005	1213	1	3.223
## 1611	85008828873	4.199	2005	1213	1	2.846
## 1623	85012550597	4.199	2005	1213	1	2.950
## 1659	85008524788	3.997	2005	1213	2	2.731
## 1662	85008541082	4.668	2005	1213	2	3.331
## 1734	43249154616	3.958	2006	1213	1	2.897
## 1736	56749181819	4.476	2006	1213	1	3.415
## 1741	61149680861	4.620	2006	1213	1	3.559
## 1743	61449316310	7.375	2006	1213	1	6.314
## 1870	33748493765	4.219	2006	1210	2	3.158
## 1879	60949256671	3.958	2006	1213	1	2.897
## 1896	51249133709	4.147	2006	1213	1	3.086
## 1909	85008814567	4.147	2006	1213	1	2.982
## 1938	60949428823	4.945	2006	1208	2	3.780
## 1952	70449761232	4.357	2006	1208	2	3.192
## 1972	85008541069	4.250	2006	1213	2	3.189
## 2060	43249137313	4.625	2007	1213	1	3.543
## 2062	51249143307	3.622	2007	1213	1	2.645
## 2074	61049412681	3.622	2007	1213	1	2.645
## 2078	61449415197	4.625	2007	1213	1	3.543
## 2079	62749095981	4.089	2007	1213	1	3.007



## 2082	63849169071	3.622	2007	1213	1	2.540
## 2089	63849342094	4.089	2007	1213	1	3.112
## 2105	60949396166	3.834	2007	1208	2	2.857
## 2131	61949171840	3.629	2007	1208	2	2.652
## 2201	34250648363	3.919	2007	1213	2	2.837
## 2220	36048982830	6.200	2007	1210	2	5.223
## 2238	51249146700	3.870	2007	1213	1	2.788
## 2242	51249158469	4.089	2007	1213	1	3.112
## 2243	62449284443	4.625	2007	1213	1	3.648
## 2245	66249153136	5.835	2007	1213	1	4.841
## 2248	77954480003	5.575	2007	1213	1	4.598
## 2256	85012549979	5.835	2007	1213	1	4.753
## 2273	61149608982	3.629	2007	1208	2	2.547
## 2420	57749138210	3.939	2008	1213	1	2.771
## 2423	61049242834	4.598	2008	1213	1	3.430
## 2431	61449387618	5.541	2008	1213	1	4.269
## 2438	63849134726	5.079	2008	1213	1	3.911
## 2454	60950457792	4.816	2008	1213	2	3.648
## 2481	57749139264	3.783	2008	1213	3	2.511
## 2521	84973622246	4.091	2008	1206	3	2.923
## 2566	57049098170	5.016	2008	1210	2	3.848
## 2615	84904577020	4.298	2008	1213	1	3.130
## 2655	70449388215	4.476	2008	1213	2	3.204
## 2814	77950377327	4.274	2009	1212	2	3.246
## 2816	77950414533	4.382	2009	1212	2	3.250
## 2850	68349152348	4.361	2009	1213	1	3.333
## 2900	84863756017	3.856	2009	1213	1	2.724
## 2919	70449753866	4.612	2009	1213	2	3.584
## 2927	77949694173	4.988	2009	1208	2	3.856
## 3042	79955423767	3.903	2010	1210	2	3.013
## 3079	77955735971	3.545	2010	1213	3	2.655
## 3103	78249271418	3.777	2010	1210	2	2.887
## 3403	84857698666	4.563	2011	1208	2	3.495
## 3429	84855915732	4.594	2011	1208	4	3.630
## 3430	84855916060	4.298	2011	1208	4	3.230
## 3431	84855927603	3.619	2011	1208	4	2.655
## 3464	83755219444	4.311	2011	1213	1	3.243
## 3699	79953040564	3.817	2011	1213	1	2.749
## 3717	79956198385	4.972	2011	1208	2	3.904
## 3722	79952723945	3.604	2011	1213	3	2.536
## 3865	84871110200	3.607	2012	1210	2	2.759
## 3867	84871243321	4.302	2012	1208	2	3.454
## 3966	84870694677	4.195	2012	1213	1	3.347
## 3967	84870697316	3.615	2012	1213	1	2.767
## 3969	84870708659	3.615	2012	1213	1	2.663
## 3973	84872981155	3.615	2012	1213	1	2.663
## 4004	84866140479	7.287	2012	1208	2	6.439
## 4013	84868655179	3.559	2012	1208	2	2.711
## 4014	84868667202	3.962	2012	1208	2	3.010
## 4125	84861079400	3.962	2012	1208	2	3.010

```

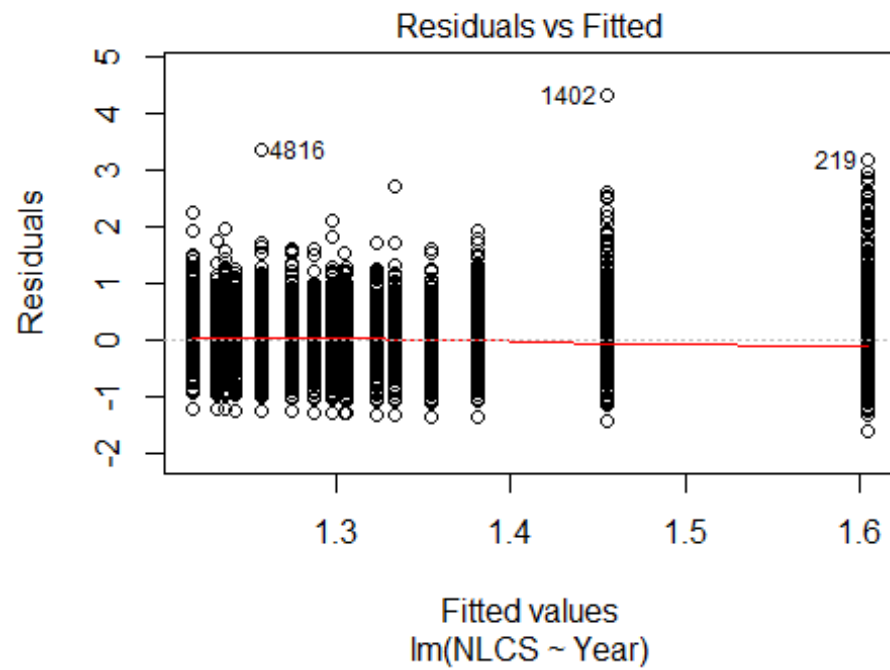
## 4134 84865872698 5.502 2012      1213      2      4.550
## 4150 84868527748 4.195 2012      1213      1      3.347
## 4154 84868545918 4.195 2012      1213      1      3.243
## 4172 84859957258 3.520 2012      1202      4      2.672
## 4182 84861566134 3.608 2012      1213      3      2.656
## 4230 84863402857 3.926 2012      1213      1      2.974
## 4234 84856425818 5.835 2012      1208      2      4.883
## 4236 84856507096 3.559 2012      1208      2      2.711
## 4307 84869102856 3.615 2012      1213      1      2.663
## 4337 84865745109 3.962 2012      1208      2      3.010
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3477 -0.8535 -0.0455  0.7199  6.4335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7679     0.0985   7.80 8.1e-15 ***
## LastAuthorFemale1  0.0958     0.0349   2.75 0.00603 **
## Year1997         -0.1987     0.1222  -1.63 0.10402
## Year1998         -0.0681     0.1237  -0.55 0.58172
## Year1999         -0.0193     0.1371  -0.14 0.88816
## Year2000         -0.0121     0.1240  -0.10 0.92254
## Year2001         -0.0145     0.1289  -0.11 0.91033
## Year2002          0.1756     0.1262   1.39 0.16409
## Year2003          0.3694     0.1358   2.72 0.00654 **
## Year2004          0.3597     0.1299   2.77 0.00563 **
## Year2005          0.4839     0.1183   4.09 4.4e-05 ***
## Year2006          0.2973     0.1178   2.52 0.01163 *
## Year2007          0.2142     0.1168   1.83 0.06664 .
## Year2008          0.4040     0.1174   3.44 0.00058 ***
## Year2009          0.2622     0.1205   2.18 0.02963 *
## Year2010          0.1226     0.1103   1.11 0.26674
## Year2011          0.2008     0.1100   1.83 0.06800 .
## Year2012          0.0855     0.1077   0.79 0.42699
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 1.01
## Multiple R-squared:  0.0278, Adjusted R-squared:  0.0233
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 7 observations c(543,1433,1834,1858,1868,3295,3479)
## are outliers with |weight| = 0 ( < 2.7e-05);

```

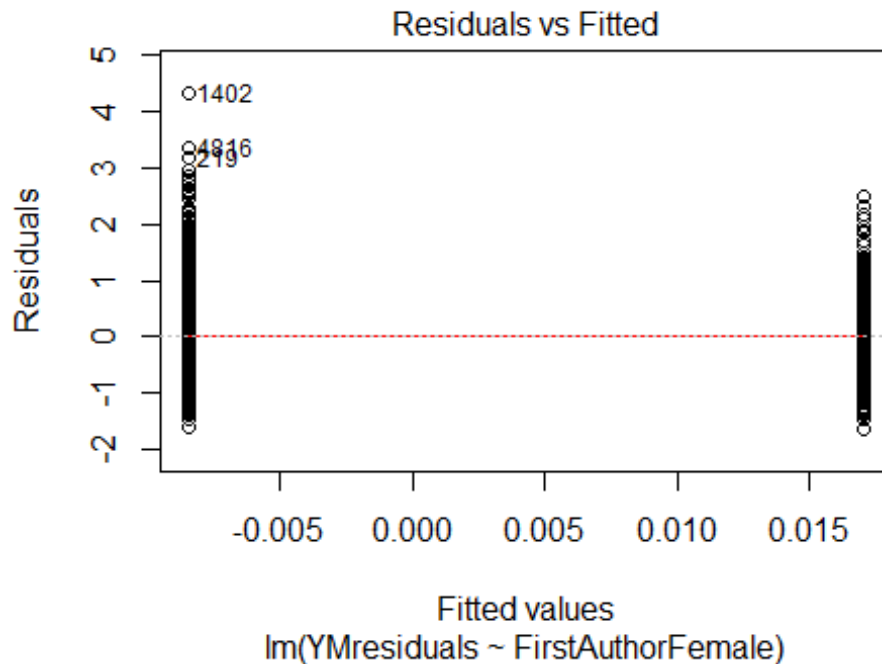
```

## 310 weights are ~= 1. The remaining 3362 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0033 0.8900 0.9360 0.9020 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3679"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1300"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 735 679 637 565 573 526 613 506 441 385 495 551 552 674 805
## 2011 2012
## 1141 1538
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 379 313 328 340 304 185 436 327 291 251 345 402 402 508 633
## 2011 2012
## 917 1233
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 354 283 306 308 279 165 384 284 252 212 299 350 353 468 570
## 2011 2012
## 821 1098
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 720, df = 16, p-value <2e-16

```

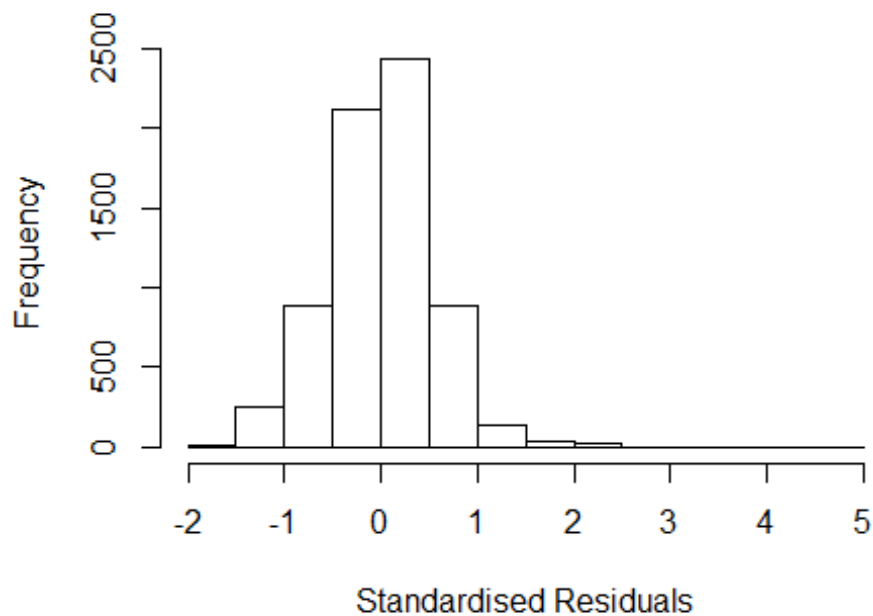


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 31, df = 1, p-value = 2e-08
```



```
## [1] "Female first author team size 2018 geometric mean: 4.49771097452171"
## [1] "Male first author team size 2018 geometric mean: 3.91793012767823"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140000, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.96179370232226"
## [1] "Male last author team size 2018 geometric mean: 4.2526157107728"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1          1.025
## LastAuthorFemale  1.032 1          1.016
## UniqueAuthors    1.168 4          1.020
## Year             1.184 16          1.005
```

## Residuals from first and last author and team size



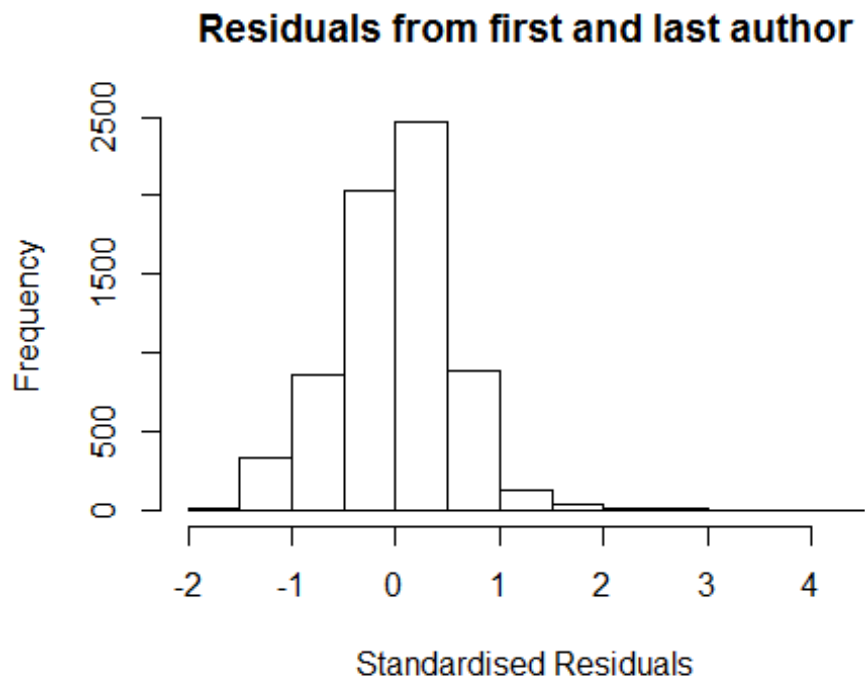
```
## [1] "List of 9 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 219    0030595341 4.756 1996    1300      1    3.234
## 222    0029796404 4.540 1996    1300      2    2.839
## 452    0030584084 4.378 1996    1300      1    2.819
## 453    0030584086 4.451 1996    1300      1    3.197
## 1369   0031471203 3.945 1997    1300      1    2.544
## 1402   33646767009 5.759 1997    1300      1    4.645
## 3040   0141452964 4.026 1999    1300      1    2.965
## 4816   0001752768 4.604 2002    1300      2    3.616
## 14257  84864217209 3.441 2012    1300      1    2.545
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7050 -0.3325  0.0216  0.3425  4.6450
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.25434    0.05631   22.28 < 2e-16 ***
## FirstAuthorFemale1 0.00437    0.01370    0.32  0.74967
```

```

## LastAuthorFemale1 -0.01571    0.01592   -0.99  0.32375
## UniqueAuthors2    0.28290    0.02515  11.25 < 2e-16 ***
## UniqueAuthors3    0.30427    0.02570  11.84 < 2e-16 ***
## UniqueAuthors4    0.33159    0.02679  12.38 < 2e-16 ***
## UniqueAuthors5    0.44631    0.02367  18.85 < 2e-16 ***
## Year1997          -0.14038    0.07418   -1.89  0.05848 .
## Year1998          -0.12076    0.07140   -1.69  0.09080 .
## Year1999          -0.19351    0.06293   -3.07  0.00211 **
## Year2000          -0.19921    0.06389   -3.12  0.00183 **
## Year2001          -0.31083    0.07547   -4.12  3.9e-05 ***
## Year2002          -0.26603    0.06093   -4.37  1.3e-05 ***
## Year2003          -0.24663    0.06365   -3.87  0.00011 ***
## Year2004          -0.25010    0.06175   -4.05  5.2e-05 ***
## Year2005          -0.24206    0.06267   -3.86  0.00011 ***
## Year2006          -0.19766    0.05963   -3.31  0.00092 ***
## Year2007          -0.23808    0.05861   -4.06  4.9e-05 ***
## Year2008          -0.22958    0.05982   -3.84  0.00013 ***
## Year2009          -0.23248    0.05859   -3.97  7.3e-05 ***
## Year2010          -0.30969    0.05812   -5.33  1.0e-07 ***
## Year2011          -0.33229    0.05624   -5.91  3.6e-09 ***
## Year2012          -0.35850    0.05571   -6.43  1.3e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.0844, Adjusted R-squared:  0.0814
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 15 observations
c(74,77,111,120,179,185,186,194,476,512,554,571,1209,1856,6147)
## are outliers with |weight| = 0 ( < 1.5e-05);
## 574 weights are ~= 1. The remaining 6197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  0.863  0.950  0.896  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.47e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"

```

```
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1          1.015
## LastAuthorFemale  1.023 1          1.011
## Year              1.050 16          1.002
```



```
## [1] "List of 15 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 219  0030595341 4.756 1996    1300      1    3.300
## 222  0029796404 4.540 1996    1300      2    3.068
## 295  0030581164 4.044 1996    1300      1    2.572
## 306  15844394511 4.140 1996    1300      2    2.668
## 327  15844394270 4.154 1996    1300      1    2.682
## 434  0030051385 4.187 1996    1300      1    2.715
## 452  0030584084 4.378 1996    1300      1    2.906
## 453  0030584086 4.451 1996    1300      1    2.979
## 474  0029952713 4.177 1996    1300      1    2.705
## 1192 0031444238 4.035 1997    1300      1    2.692
## 1273 0030911052 3.980 1997    1300      1    2.637
## 1369 0031471203 3.945 1997    1300      1    2.559
## 1402 33646767009 5.759 1997    1300      1    4.416
## 3040 0141452964 4.026 1999    1300      1    2.706
## 4816 0001752768 4.604 2002    1300      2    3.344
##
## Call:
```

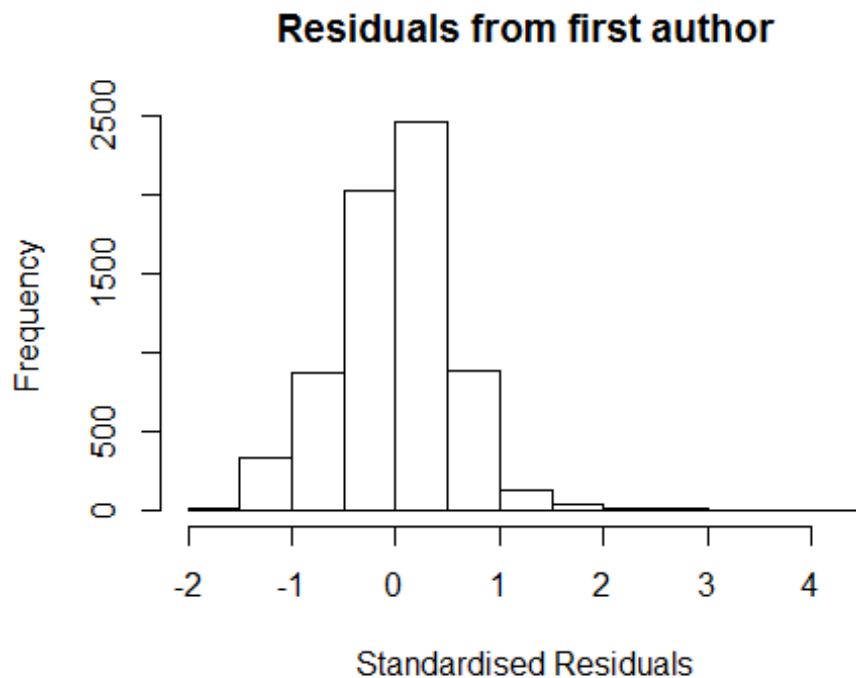


```

## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.5161 -0.3481  0.0299  0.3489  4.4162
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4724     0.0514   28.67 < 2e-16 ***
## FirstAuthorFemale1  0.0437     0.0140    3.13  0.00176 **
## LastAuthorFemale1 -0.0160     0.0166   -0.97  0.33378
## Year1997          -0.1296     0.0725   -1.79  0.07399 .
## Year1998          -0.1145     0.0690   -1.66  0.09727 .
## Year1999          -0.1519     0.0610   -2.49  0.01273 *
## Year2000          -0.1466     0.0616   -2.38  0.01735 *
## Year2001          -0.2346     0.0727   -3.23  0.00126 **
## Year2002          -0.2124     0.0591   -3.60  0.00033 ***
## Year2003          -0.1842     0.0618   -2.98  0.00288 **
## Year2004          -0.1899     0.0593   -3.20  0.00136 **
## Year2005          -0.1724     0.0615   -2.81  0.00503 **
## Year2006          -0.1393     0.0582   -2.40  0.01662 *
## Year2007          -0.1630     0.0562   -2.90  0.00372 **
## Year2008          -0.1470     0.0584   -2.52  0.01182 *
## Year2009          -0.1727     0.0570   -3.03  0.00247 **
## Year2010          -0.2275     0.0562   -4.05  5.2e-05 ***
## Year2011          -0.2333     0.0540   -4.32  1.6e-05 ***
## Year2012          -0.2611     0.0533   -4.90  1.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.0109
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 17 observations
## c(16,74,77,106,110,111,120,179,185,186,194,476,512,554,571,1209,1856)
## are outliers with |weight| = 0 ( < 1.5e-05);
## 543 weights are ~ 1. The remaining 6226 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0003  0.8630  0.9510  0.8940  0.9850  0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01

```

```
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##          500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##          0          1000          0
##          psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
##      compute.outlier.stats
##          "SM"
##      seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1          1.014
## Year              1.028 16          1.001
```



```
## [1] "List of 15 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 219  0030595341 4.756 1996    1300      1    3.300
## 222  0029796404 4.540 1996    1300      2    3.068
## 295  0030581164 4.044 1996    1300      1    2.572
## 306  15844394511 4.140 1996    1300      2    2.668
## 327  15844394270 4.154 1996    1300      1    2.682
## 434  0030051385 4.187 1996    1300      1    2.715
## 452  0030584084 4.378 1996    1300      1    2.906
## 453  0030584086 4.451 1996    1300      1    2.979
## 474  0029952713 4.177 1996    1300      1    2.705
## 1192 0031444238 4.035 1997    1300      1    2.692
## 1273 0030911052 3.980 1997    1300      1    2.637
```

```

## 1369 0031471203 3.945 1997 1300 1 2.559
## 1402 33646767009 5.759 1997 1300 1 4.416
## 3040 0141452964 4.026 1999 1300 1 2.706
## 4816 0001752768 4.604 2002 1300 2 3.344
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5120 -0.3477  0.0304  0.3486  4.4193
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4702     0.0513   28.67 < 2e-16 ***
## FirstAuthorFemale1  0.0418     0.0140    2.98  0.00290 **
## Year1997          -0.1305     0.0725   -1.80  0.07213 .
## Year1998          -0.1144     0.0690   -1.66  0.09751 .
## Year1999          -0.1516     0.0610   -2.48  0.01304 *
## Year2000          -0.1468     0.0616   -2.38  0.01724 *
## Year2001          -0.2343     0.0727   -3.22  0.00128 **
## Year2002          -0.2115     0.0591   -3.58  0.00035 ***
## Year2003          -0.1841     0.0618   -2.98  0.00290 **
## Year2004          -0.1895     0.0593   -3.20  0.00140 **
## Year2005          -0.1722     0.0614   -2.80  0.00508 **
## Year2006          -0.1396     0.0582   -2.40  0.01650 *
## Year2007          -0.1633     0.0562   -2.91  0.00368 **
## Year2008          -0.1470     0.0584   -2.52  0.01193 *
## Year2009          -0.1740     0.0571   -3.05  0.00231 **
## Year2010          -0.2278     0.0563   -4.05  5.2e-05 ***
## Year2011          -0.2340     0.0540   -4.33  1.5e-05 ***
## Year2012          -0.2625     0.0534   -4.92  8.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0134, Adjusted R-squared:  0.0109
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 17 observations
c(16,74,77,106,110,111,120,179,185,186,194,476,512,554,571,1209,1856)
## are outliers with |weight| = 0 ( < 1.5e-05);
## 546 weights are ~ 1. The remaining 6223 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0002  0.8630  0.9510  0.8940  0.9850  0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol

```

```

##          1.55e+00          5.00e-01          4.69e+00          1.00e-07
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          1.47e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##          500          50          2          1          1000          200
## trace.lev mts compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1          1.010
## Year          1.021 16          1.001

## [1] "List of 15 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 219 0030595341 4.756 1996 1300 1 3.300
## 222 0029796404 4.540 1996 1300 2 3.068
## 295 0030581164 4.044 1996 1300 1 2.572
## 306 15844394511 4.140 1996 1300 2 2.668
## 327 15844394270 4.154 1996 1300 1 2.682
## 434 0030051385 4.187 1996 1300 1 2.715
## 452 0030584084 4.378 1996 1300 1 2.906
## 453 0030584086 4.451 1996 1300 1 2.979
## 474 0029952713 4.177 1996 1300 1 2.705
## 1192 0031444238 4.035 1997 1300 1 2.692
## 1273 0030911052 3.980 1997 1300 1 2.637
## 1369 0031471203 3.945 1997 1300 1 2.559
## 1402 33646767009 5.759 1997 1300 1 4.416
## 3040 0141452964 4.026 1999 1300 1 2.706
## 4816 0001752768 4.604 2002 1300 2 3.344
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##          Min          1Q        Median          3Q          Max
## -1.4796 -0.3495  0.0318  0.3498  4.4080
##
## Coefficients:
##          Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.47964    0.05150   28.73 < 2e-16 ***
## LastAuthorFemale1 -0.00882    0.01657   -0.53  0.59466

```

```

## Year1997      -0.12864      0.07276      -1.77      0.07711      .
## Year1998      -0.11150      0.06917      -1.61      0.10698
## Year1999      -0.14872      0.06118      -2.43      0.01509      *
## Year2000      -0.14168      0.06172      -2.30      0.02174      *
## Year2001      -0.23321      0.07292      -3.20      0.00139      **
## Year2002      -0.20955      0.05929      -3.53      0.00041      ***
## Year2003      -0.17935      0.06182      -2.90      0.00373      **
## Year2004      -0.18450      0.05941      -3.11      0.00191      **
## Year2005      -0.16829      0.06160      -2.73      0.00631      **
## Year2006      -0.13350      0.05824      -2.29      0.02194      *
## Year2007      -0.15758      0.05626      -2.80      0.00511      **
## Year2008      -0.14119      0.05848      -2.41      0.01579      *
## Year2009      -0.16694      0.05717      -2.92      0.00351      **
## Year2010      -0.21904      0.05629      -3.89      0.00010      ***
## Year2011      -0.22432      0.05397      -4.16      3.3e-05      ***
## Year2012      -0.25169      0.05334      -4.72      2.4e-06      ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0121, Adjusted R-squared:  0.00958
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 17 observations
## c(16,74,77,106,110,111,120,179,185,186,194,476,512,554,571,1209,1856)
## are outliers with |weight| = 0 ( < 1.5e-05);
## 538 weights are ~ 1. The remaining 6231 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0006 0.8640 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
## tuning.chi      bb      tuning.psi      refine.tol
## 1.55e+00      5.00e-01      4.69e+00      1.00e-07
## rel.tol      solve.tol      eps.outlier      eps.x
## 1.00e-07      1.00e-07      1.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
## 500      50      2      1      1000      200
## trace.lev      mts      compute.rd
## 0      1000      0
## psi      subsampling      cov
## "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6786"
## [1] ""
## [1] ""
## [1] "#####"

```

```

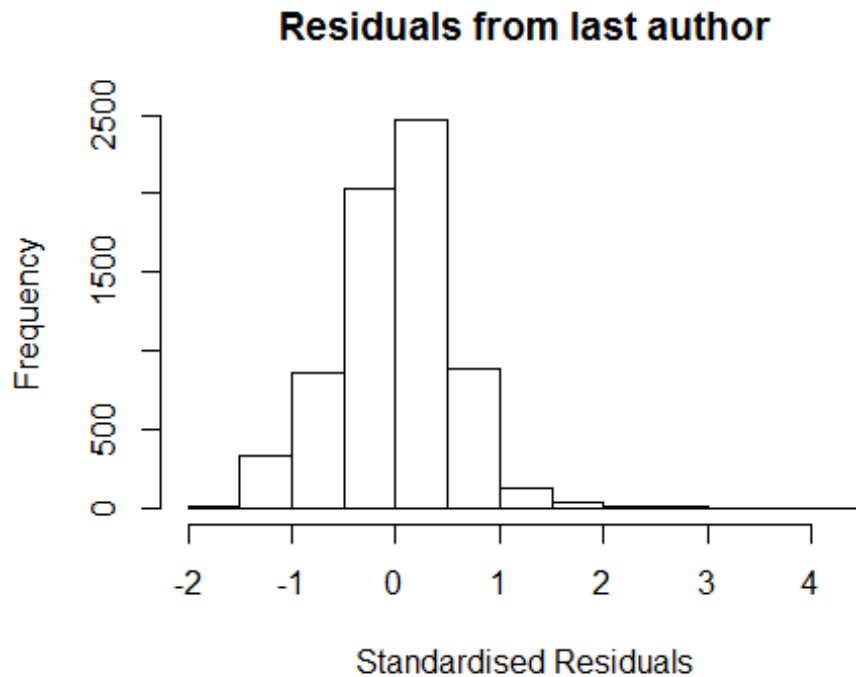
## [1] "Analysis of AJSC 1301"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2007 2008 2009 2010 2011 2012
##    1    5    4    2    2   11
##
## 2007 2008 2009 2010 2011 2012
##    0    5    3    1    2    7
##
## 2007 2008 2009 2010 2011 2012
##    0    5    2    1    2    6
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.34006081041922"
## [1] "Male first author team size 2018 geometric mean: 3.88769612712021"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.85993829661746"
## [1] "Male last author team size 2018 geometric mean: 4.125125285755"

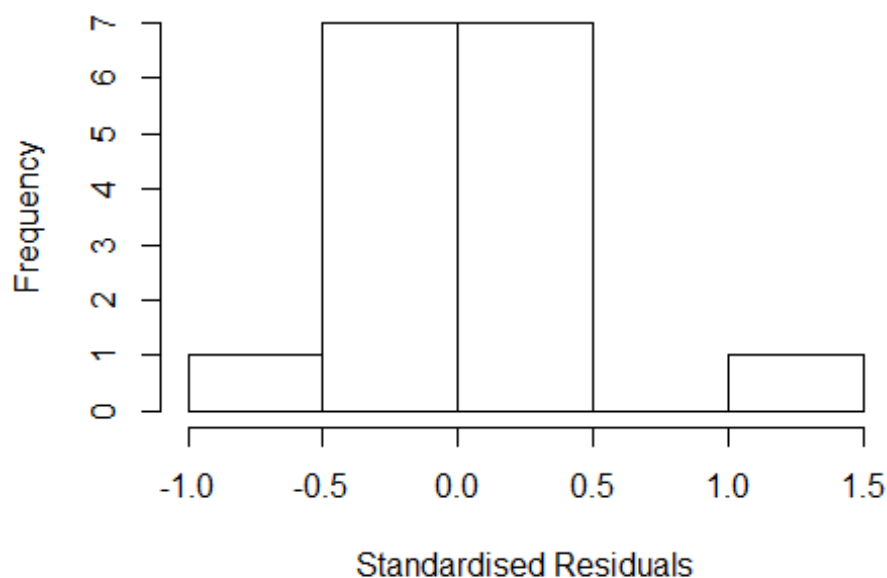
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
##      Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
##      factors"
##               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 586.1 1         24.209
## LastAuthorFemale  201.4 1         14.190
## Year              8192.6 4          3.084
```

## Residuals from first and last author

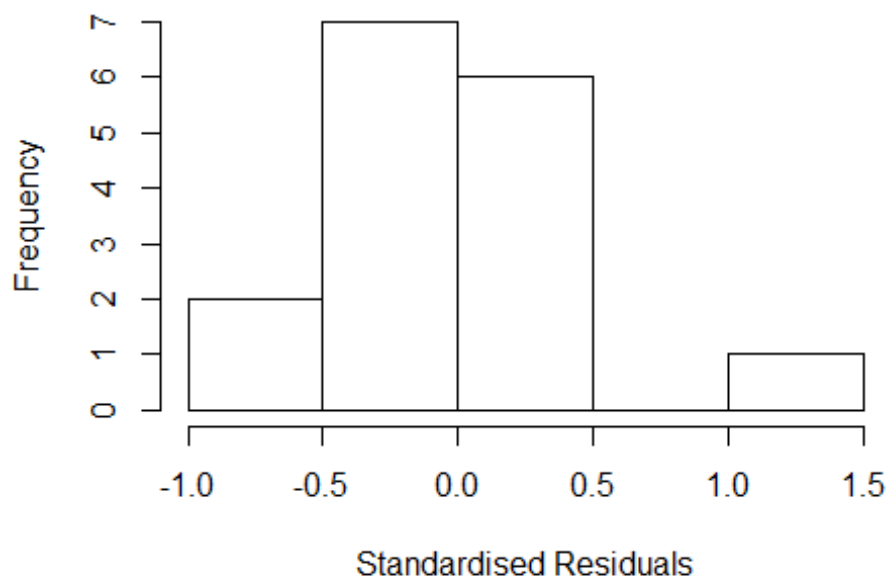


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.56616 -0.24675 0.00465 0.14148 1.29784
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.431 0.101 4.29 0.002 **
## FirstAuthorFemale1 0.971 0.402 2.41 0.039 *
## LastAuthorFemale1 0.598 0.252 2.38 0.041 *
## Year2009 0.540 0.240 2.25 0.051 .
## Year2010 0.134 0.404 0.33 0.748
## Year2011 0.534 0.257 2.08 0.067 .
## Year2012 0.135 0.252 0.54 0.605
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared: 0.434, Adjusted R-squared: 0.0571
```



```
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 12 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.506  0.923  0.969   0.925   0.985   0.997
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.25e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 14.44  1      3.799
## Year              14.44  4      1.396
```

### Residuals from first author

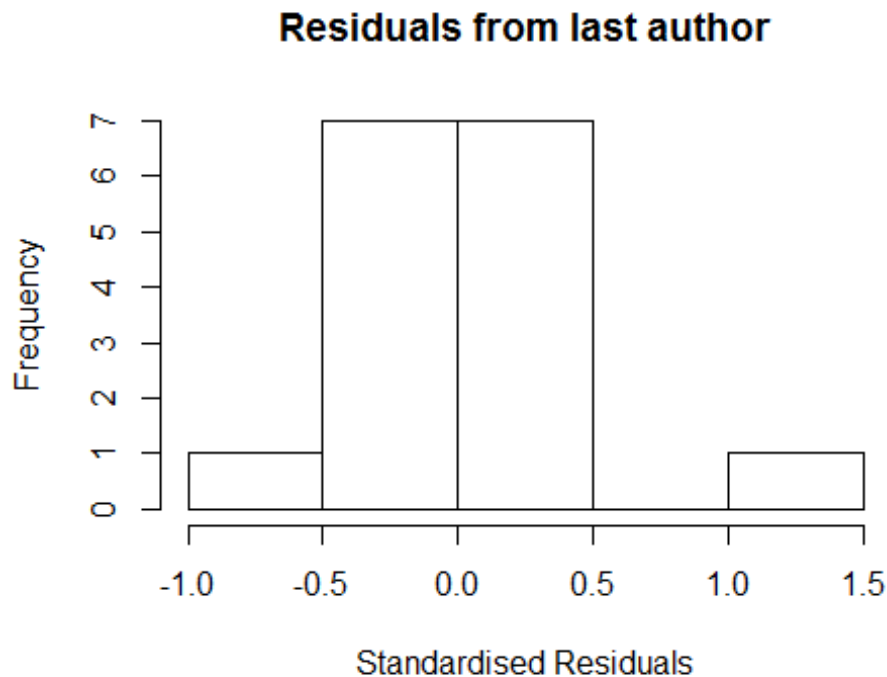


```

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.8537 -0.3017 -0.0527 0.3194 1.0103
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.628 0.177 3.55 0.0052 **
## FirstAuthorFemale1 0.774 0.406 1.90 0.0861 .
## Year2009 0.343 0.276 1.25 0.2415
## Year2010 0.134 0.366 0.37 0.7218
## Year2011 0.337 0.288 1.17 0.2693
## Year2012 0.225 0.321 0.70 0.4991
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.746
## Multiple R-squared: 0.229, Adjusted R-squared: -0.156
## Convergence in 8 IRWLS iterations
##
## Robustness weights:
## one weight is ~= 1. The remaining 15 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.840 0.963 0.984 0.964 0.986 0.998
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 6.25e-03 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 1000 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2*Df))

```

```
## LastAuthorFemale 19.66 1 4.434
## Year 19.66 4 1.451
```



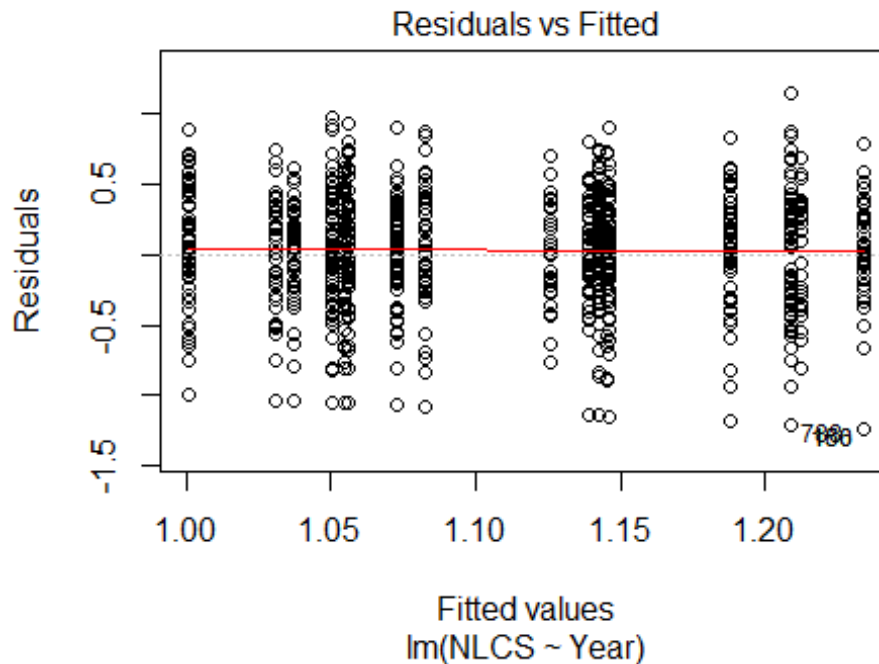
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.6584 -0.2957 -0.0027 0.1729 1.2056
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.790 0.314 2.52 0.031 *
## LastAuthorFemale1 0.418 0.316 1.32 0.215
## Year2009 0.181 0.382 0.47 0.645
## Year2010 0.746 0.314 2.37 0.039 *
## Year2011 0.175 0.392 0.45 0.664
## Year2012 -0.132 0.345 -0.38 0.711
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## Robust residual standard error: 0.543
## Multiple R-squared: 0.193, Adjusted R-squared: -0.211
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 3 weights are ~= 1. The remaining 13 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.601  0.936  0.970   0.915  0.978   0.995
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           6.25e-03           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 16"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1302"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   65   60   79   74   70  107  107   73   79   97   94   86   82  109   81
## 2011 2012
##   89  128
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   29   47   38   33   47   76   43   57   69   58   63   54   89   62
## 2011 2012
##   66  104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   20   25   39   34   30   42   69   35   49   59   50   57   49   85   60
## 2011 2012
##   59   95
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##

```

```
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



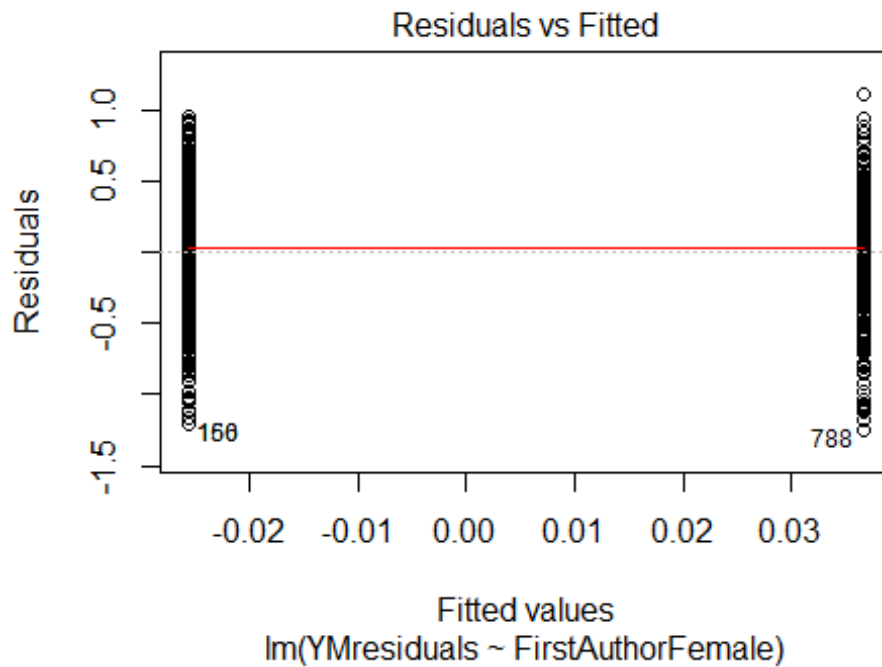
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.1, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 4.46270538468545"
## [1] "Male first author team size 2018 geometric mean: 4.70134621868663"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

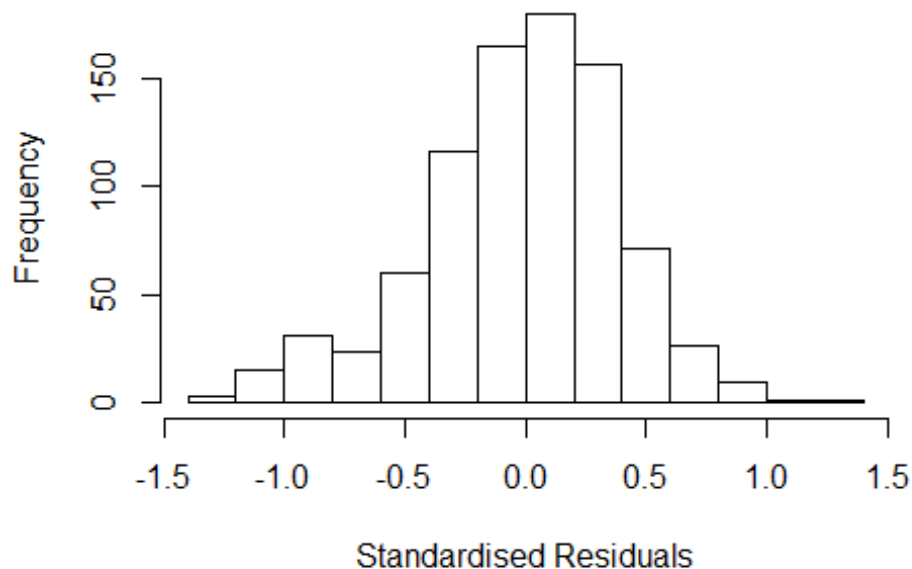
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 390, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.86322295685879"
## [1] "Male last author team size 2018 geometric mean: 4.39847734369168"
```

```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.304 1      1.142
## LastAuthorFemale  1.185 1      1.089
## UniqueAuthors    1.566 4      1.058
## Year              1.990 16     1.022
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3071 -0.2468 0.0192 0.2624 1.2924
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01872 0.11667 8.73 < 2e-16 ***
## FirstAuthorFemale1 0.02768 0.03011 0.92 0.35820
## LastAuthorFemale1 0.07426 0.03209 2.31 0.02090 *
## UniqueAuthors2 0.09132 0.06032 1.51 0.13041
## UniqueAuthors3 0.19849 0.05992 3.31 0.00096 ***
## UniqueAuthors4 0.19981 0.05942 3.36 0.00081 ***
## UniqueAuthors5 0.27967 0.05266 5.31 1.4e-07 ***
## Year1997 0.03202 0.14218 0.23 0.82189
## Year1998 0.00875 0.12120 0.07 0.94249
## Year1999 -0.17132 0.12967 -1.32 0.18681
```

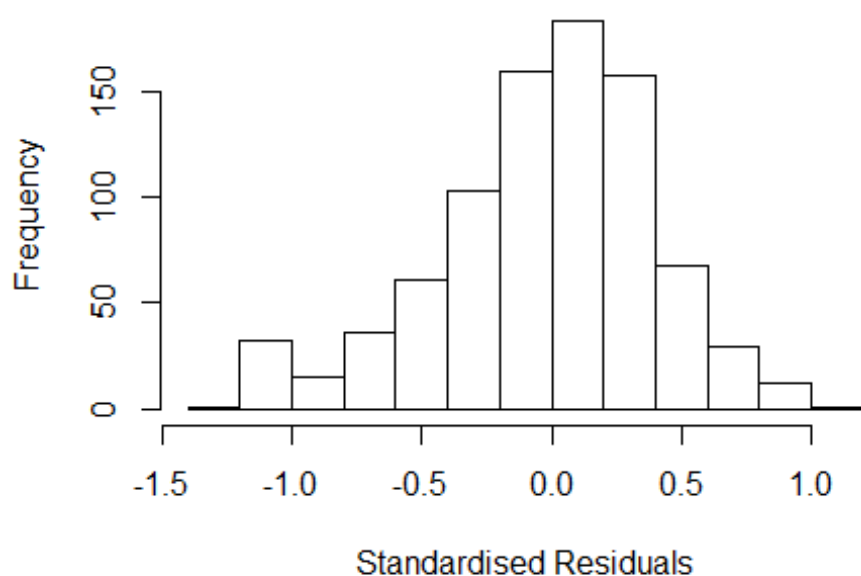
```

## Year2000      -0.12027    0.12860   -0.94  0.34992
## Year2001      -0.15071    0.13312   -1.13  0.25788
## Year2002      -0.13824    0.12007   -1.15  0.24993
## Year2003      -0.06423    0.12250   -0.52  0.60021
## Year2004      -0.07408    0.13199   -0.56  0.57476
## Year2005      -0.19517    0.12632   -1.54  0.12273
## Year2006      -0.00456    0.13052   -0.03  0.97211
## Year2007      -0.15035    0.13108   -1.15  0.25169
## Year2008      -0.16670    0.12067   -1.38  0.16748
## Year2009      -0.13231    0.11705   -1.13  0.25867
## Year2010      -0.16749    0.11899   -1.41  0.15961
## Year2011      -0.09681    0.11855   -0.82  0.41439
## Year2012      -0.08824    0.11511   -0.77  0.44353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0857, Adjusted R-squared:  0.0615
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 783 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.214  0.870  0.952  0.895  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.197 1      1.094
## LastAuthorFemale  1.146 1      1.071
## Year              1.351 16      1.009

```



## Residuals from first and last author



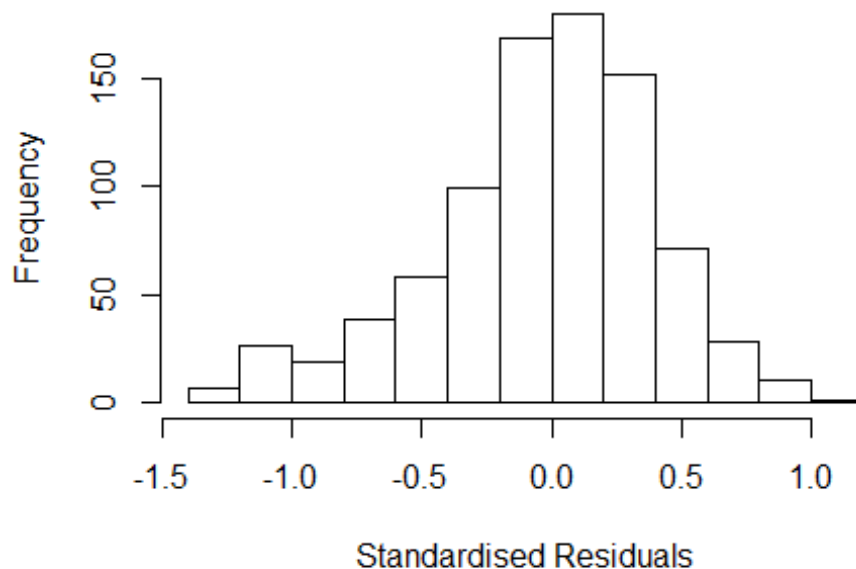
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2510 -0.2612 0.0183 0.2597 1.1829
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11881 0.10873 10.29 <2e-16 ***
## FirstAuthorFemale1 0.04610 0.02976 1.55 0.122
## LastAuthorFemale1 0.07793 0.03241 2.40 0.016 *
## Year1997 0.07314 0.13835 0.53 0.597
## Year1998 0.07839 0.12036 0.65 0.515
## Year1999 -0.12017 0.12660 -0.95 0.343
## Year2000 -0.03320 0.12793 -0.26 0.795
## Year2001 -0.05407 0.13357 -0.40 0.686
## Year2002 -0.11534 0.12026 -0.96 0.338
## Year2003 0.02157 0.12378 0.17 0.862
## Year2004 0.00814 0.13595 0.06 0.952
## Year2005 -0.11303 0.12752 -0.89 0.376
```

```

## Year2006          0.06503      0.12994      0.50      0.617
## Year2007          -0.07142      0.13023     -0.55      0.584
## Year2008          -0.08758      0.12068     -0.73      0.468
## Year2009          -0.06537      0.11571     -0.56      0.572
## Year2010          -0.08041      0.12014     -0.67      0.503
## Year2011          -0.00257      0.11842     -0.02      0.983
## Year2012           0.01062      0.11519      0.09      0.927
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.0342, Adjusted R-squared:  0.0135
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 792 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.265  0.871  0.950  0.893  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.21 1      1.100
## Year              1.21 16      1.006

```

## Residuals from first author



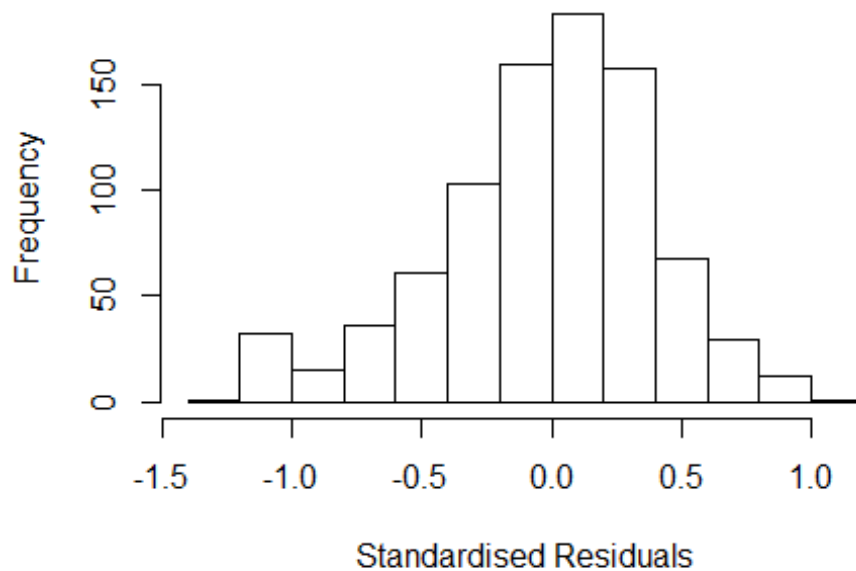
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2056 -0.2594 0.0218 0.2694 1.1545
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1256 0.1109 10.15 <2e-16 ***
## FirstAuthorFemale1 0.0536 0.0300 1.79 0.075 .
## Year1997 0.0682 0.1415 0.48 0.630
## Year1998 0.0800 0.1219 0.66 0.512
## Year1999 -0.1175 0.1277 -0.92 0.357
## Year2000 -0.0174 0.1304 -0.13 0.894
## Year2001 -0.0456 0.1360 -0.34 0.738
## Year2002 -0.1092 0.1222 -0.89 0.372
## Year2003 0.0219 0.1255 0.17 0.861
## Year2004 0.0223 0.1377 0.16 0.871
## Year2005 -0.1012 0.1293 -0.78 0.434
## Year2006 0.0761 0.1303 0.58 0.559
```

```

## Year2007          -0.0570      0.1308   -0.44    0.663
## Year2008          -0.0754      0.1218   -0.62    0.536
## Year2009          -0.0504      0.1172   -0.43    0.667
## Year2010          -0.0616      0.1209   -0.51    0.610
## Year2011           0.0242      0.1190    0.20    0.839
## Year2012           0.0220      0.1168    0.19    0.851
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0271, Adjusted R-squared:  0.00743
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 789 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.870  0.948  0.893  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.157 1      1.076
## Year              1.157 16      1.005

```

## Residuals from last author



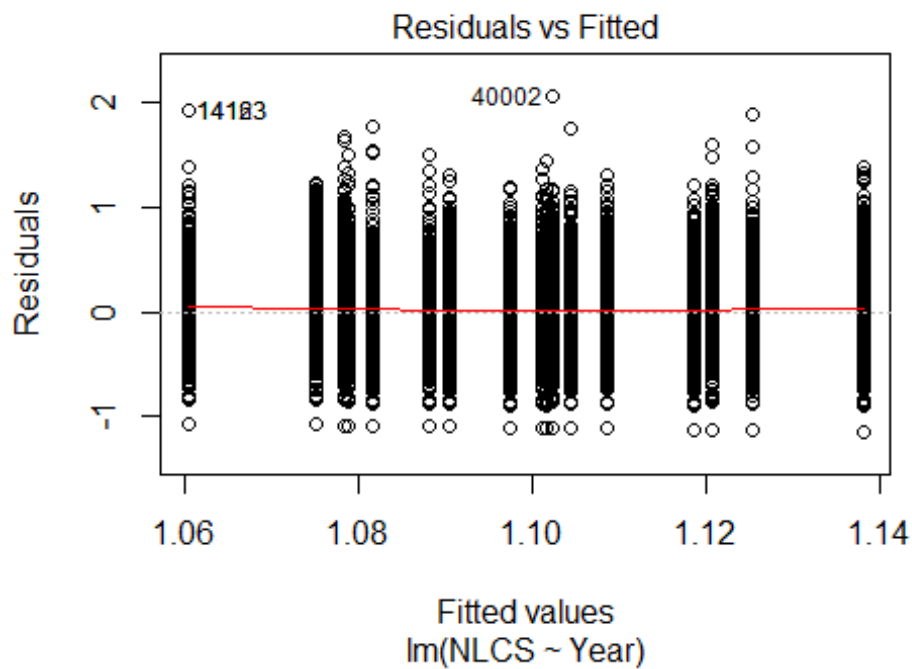
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2241 -0.2538 0.0217 0.2637 1.2152
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13425 0.10773 10.53 <2e-16 ***
## LastAuthorFemale1 0.08332 0.03259 2.56 0.011 *
## Year1997 0.07356 0.13880 0.53 0.596
## Year1998 0.07552 0.11989 0.63 0.529
## Year1999 -0.11699 0.12642 -0.93 0.355
## Year2000 -0.02726 0.12817 -0.21 0.832
## Year2001 -0.06392 0.13264 -0.48 0.630
## Year2002 -0.11850 0.12039 -0.98 0.325
## Year2003 0.02525 0.12315 0.21 0.838
## Year2004 0.00657 0.13550 0.05 0.961
## Year2005 -0.11288 0.12708 -0.89 0.375
## Year2006 0.07046 0.12850 0.55 0.584
```

```

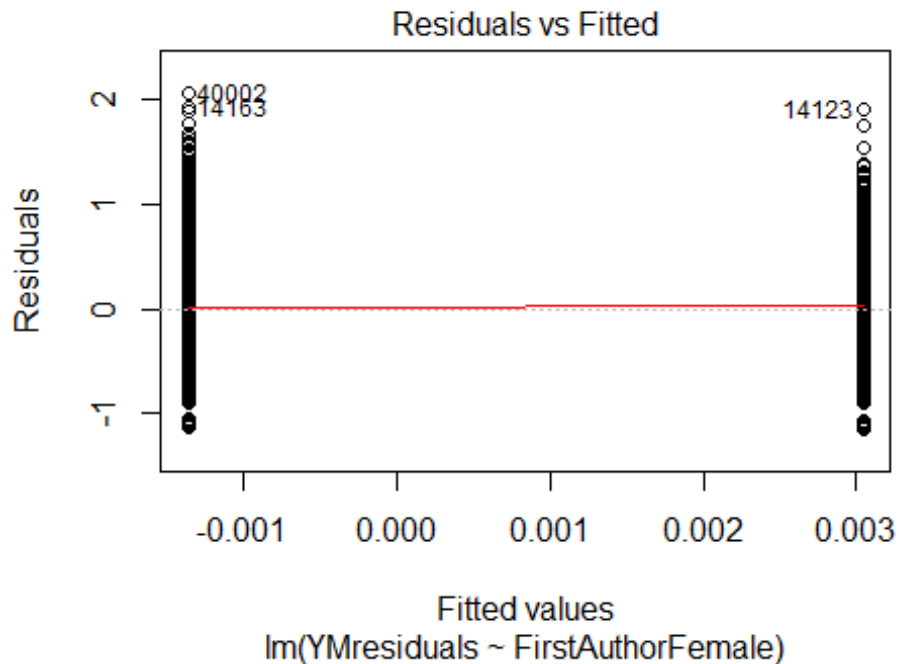
## Year2007          -0.06854      0.13058    -0.52      0.600
## Year2008          -0.08631      0.12057    -0.72      0.474
## Year2009          -0.06062      0.11545    -0.53      0.600
## Year2010          -0.07581      0.11984    -0.63      0.527
## Year2011           0.00565      0.11804      0.05      0.962
## Year2012           0.02202      0.11449      0.19      0.848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.0318, Adjusted R-squared:  0.0122
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 788 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.285  0.866  0.951  0.892  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 857"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1303"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2480 2510 2282 1996 2162 1987 2022 1917 1822 1822 1766 1659 1778 1872 1755
## 2011 2012
## 1770 1585
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1547 1592 1544 1330 1281 1090 1456 1404 1384 1268 1261 1278 1316 1434 1320
## 2011 2012

```

```
## 1379 1192
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1369 1401 1383 1185 1151 960 1314 1248 1230 1135 1133 1162 1183 1300 1149
## 2011 2012
## 1231 1070
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 970, df = 16, p-value <2e-16
```



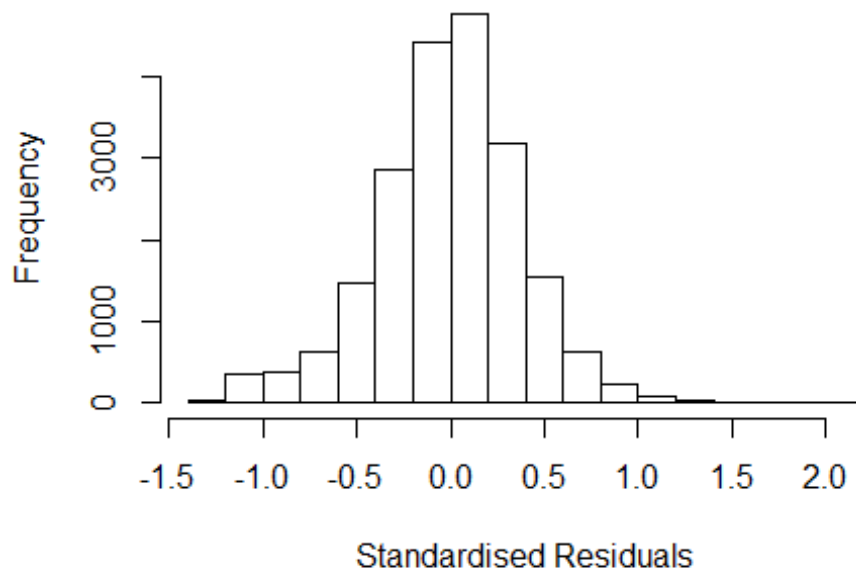
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 23, df = 1, p-value = 2e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.83232625414465"
## [1] "Male first author team size 2018 geometric mean: 4.25816193589526"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.84553834801442"
## [1] "Male last author team size 2018 geometric mean: 4.372279860374"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 31000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1          1.014
## LastAuthorFemale  1.010 1          1.005
## UniqueAuthors     1.072 4          1.009
## Year               1.087 16         1.003
```



## Residuals from first and last author and team size



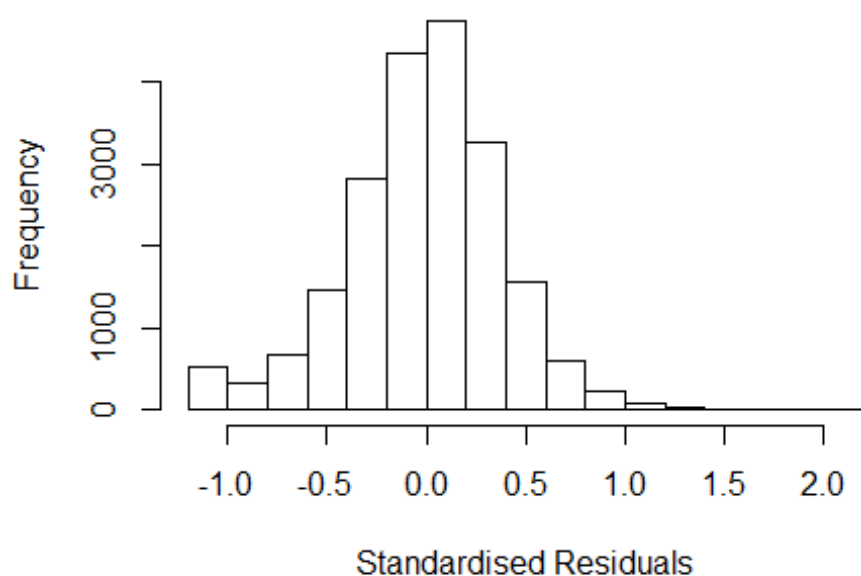
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26138 -0.23143 0.00588 0.22799 2.08319
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99827 0.02344 42.60 < 2e-16 ***
## FirstAuthorFemale1 0.00331 0.00550 0.60 0.54667
## LastAuthorFemale1 -0.00880 0.00725 -1.21 0.22505
## UniqueAuthors2 0.15538 0.02053 7.57 4.0e-14 ***
## UniqueAuthors3 0.16292 0.02008 8.12 5.1e-16 ***
## UniqueAuthors4 0.18873 0.02013 9.38 < 2e-16 ***
## UniqueAuthors5 0.25979 0.01966 13.21 < 2e-16 ***
## Year1997 -0.03681 0.01957 -1.88 0.05996 .
## Year1998 -0.06857 0.01821 -3.77 0.00017 ***
## Year1999 -0.07831 0.01670 -4.69 2.8e-06 ***
```

```

## Year2000      -0.10145    0.01738    -5.84    5.4e-09 ***
## Year2001      -0.07720    0.01733    -4.46    8.4e-06 ***
## Year2002      -0.08052    0.01590    -5.07    4.1e-07 ***
## Year2003      -0.09889    0.01596    -6.20    5.9e-10 ***
## Year2004      -0.11648    0.01583    -7.36    1.9e-13 ***
## Year2005      -0.11181    0.01629    -6.86    6.9e-12 ***
## Year2006      -0.11964    0.01645    -7.27    3.7e-13 ***
## Year2007      -0.10263    0.01656    -6.20    5.8e-10 ***
## Year2008      -0.06982    0.01661    -4.20    2.6e-05 ***
## Year2009      -0.09381    0.01643    -5.71    1.1e-08 ***
## Year2010      -0.07439    0.01686    -4.41    1.0e-05 ***
## Year2011      -0.08084    0.01721    -4.70    2.7e-06 ***
## Year2012      -0.06797    0.01798    -3.78    0.00016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.341
## Multiple R-squared:  0.035, Adjusted R-squared:  0.0339
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 9 observations c(2880,4814,6464,6480,10467,12741,17188,17718,18961)
## are outliers with |weight| = 0 ( < 4.9e-06);
## 1770 weights are ~= 1. The remaining 18825 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0019 0.8600 0.9500 0.8870 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      4.85e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1 1.009
## LastAuthorFemale 1.007 1 1.003
## Year 1.024 16 1.001

```

## Residuals from first and last author



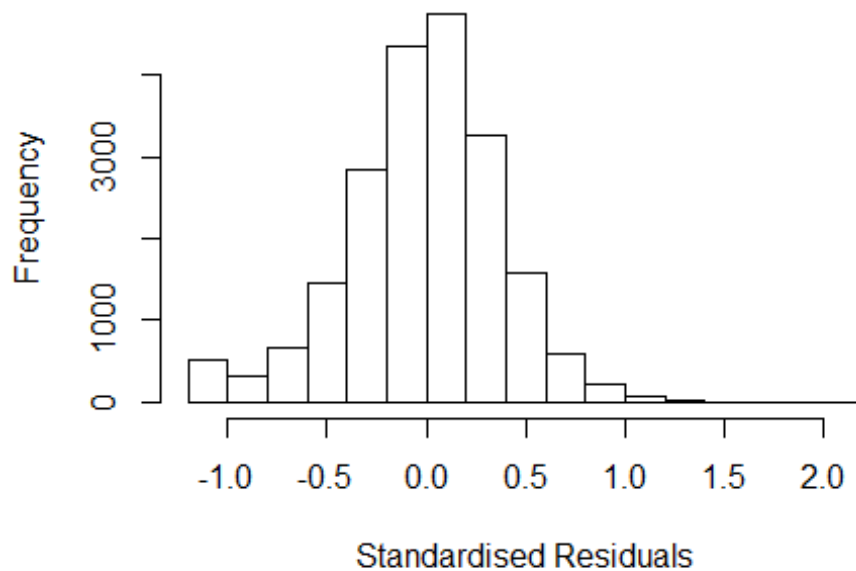
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19222 -0.23542 0.00683 0.22917 2.03585
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17957 0.01355 87.04 < 2e-16 ***
## FirstAuthorFemale1 0.01265 0.00552 2.29 0.02204 *
## LastAuthorFemale1 -0.00579 0.00730 -0.79 0.42715
## Year1997 -0.03840 0.01986 -1.93 0.05323 .
## Year1998 -0.06931 0.01843 -3.76 0.00017 ***
## Year1999 -0.07415 0.01689 -4.39 1.1e-05 ***
## Year2000 -0.10049 0.01759 -5.71 1.1e-08 ***
## Year2001 -0.06800 0.01761 -3.86 0.00011 ***
## Year2002 -0.06560 0.01617 -4.06 5.0e-05 ***
## Year2003 -0.08420 0.01624 -5.18 2.2e-07 ***
## Year2004 -0.09890 0.01606 -6.16 7.5e-10 ***
## Year2005 -0.09237 0.01645 -5.62 2.0e-08 ***
```

```

## Year2006          -0.09858      0.01672      -5.89      3.8e-09 ***
## Year2007          -0.08120      0.01672      -4.86      1.2e-06 ***
## Year2008          -0.04981      0.01681      -2.96      0.00304 **
## Year2009          -0.07410      0.01657      -4.47      7.8e-06 ***
## Year2010          -0.05155      0.01713      -3.01      0.00261 **
## Year2011          -0.05942      0.01741      -3.41      0.00065 ***
## Year2012          -0.04354      0.01838      -2.37      0.01787 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.00524,    Adjusted R-squared:  0.00437
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 7 observations c(2881,4814,6464,6480,10467,17718,18961)
## are outliers with |weight| = 0 ( < 4.9e-06);
## 1834 weights are ~= 1. The remaining 18763 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0009 0.8590 0.9500 0.8860 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.85e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1          1.009
## Year              1.018 16          1.001

```

## Residuals from first author



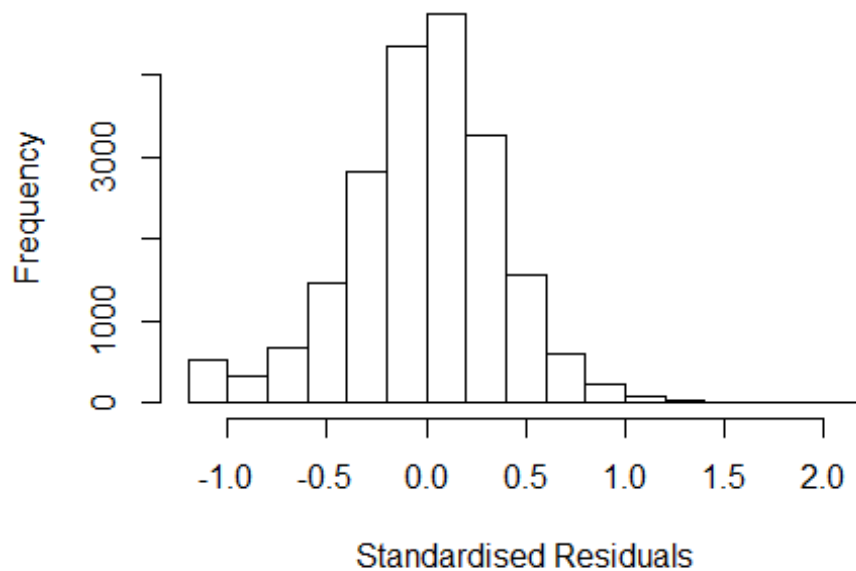
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19134 -0.23563 0.00712 0.22942 2.03674
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17888 0.01351 87.24 < 2e-16 ***
## FirstAuthorFemale1 0.01247 0.00553 2.26 0.02413 *
## Year1997 -0.03827 0.01986 -1.93 0.05391 .
## Year1998 -0.06930 0.01843 -3.76 0.00017 ***
## Year1999 -0.07423 0.01689 -4.39 1.1e-05 ***
## Year2000 -0.10058 0.01759 -5.72 1.1e-08 ***
## Year2001 -0.06813 0.01761 -3.87 0.00011 ***
## Year2002 -0.06580 0.01618 -4.07 4.8e-05 ***
## Year2003 -0.08430 0.01624 -5.19 2.1e-07 ***
## Year2004 -0.09896 0.01606 -6.16 7.4e-10 ***
## Year2005 -0.09246 0.01645 -5.62 1.9e-08 ***
## Year2006 -0.09863 0.01673 -5.90 3.8e-09 ***
```

```

## Year2007          -0.08135      0.01672      -4.87  1.1e-06 ***
## Year2008          -0.05006      0.01681      -2.98  0.00290 **
## Year2009          -0.07427      0.01658      -4.48  7.5e-06 ***
## Year2010          -0.05173      0.01712      -3.02  0.00252 **
## Year2011          -0.05962      0.01741      -3.42  0.00062 ***
## Year2012          -0.04372      0.01838      -2.38  0.01742 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.00521,    Adjusted R-squared:  0.00439
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 7 observations c(2881,4814,6464,6480,10467,17718,18961)
## are outliers with |weight| = 0 ( < 4.9e-06);
## 1820 weights are ~1. The remaining 18777 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8600 0.9510 0.8860 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.85e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.007 1          1.003
## Year            1.007 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18257 -0.23438 0.00656 0.22970 2.03127
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18257 0.01353 87.43 < 2e-16 ***
## LastAuthorFemale1 -0.00513 0.00731 -0.70 0.48316
## Year1997 -0.03834 0.01988 -1.93 0.05385 .
## Year1998 -0.06924 0.01844 -3.76 0.00017 ***
## Year1999 -0.07314 0.01688 -4.33 1.5e-05 ***
## Year2000 -0.10020 0.01760 -5.69 1.3e-08 ***
## Year2001 -0.06728 0.01762 -3.82 0.00013 ***
## Year2002 -0.06480 0.01618 -4.00 6.3e-05 ***
## Year2003 -0.08325 0.01624 -5.13 3.0e-07 ***
## Year2004 -0.09775 0.01605 -6.09 1.2e-09 ***
## Year2005 -0.09132 0.01645 -5.55 2.8e-08 ***
## Year2006 -0.09763 0.01672 -5.84 5.4e-09 ***
```

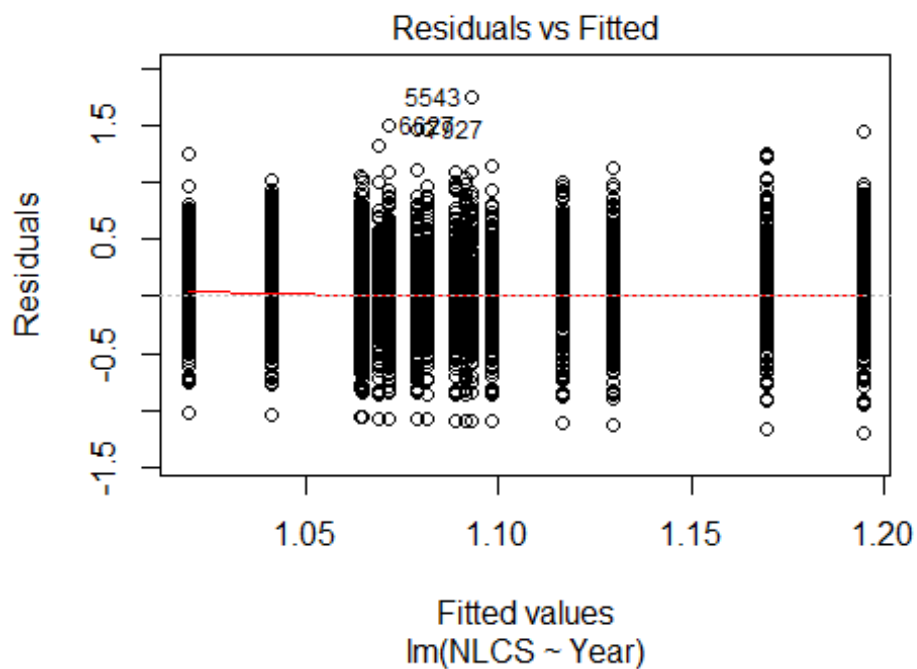
```

## Year2007          -0.08012      0.01672    -4.79  1.7e-06 ***
## Year2008          -0.04857      0.01679    -2.89  0.00383 **
## Year2009          -0.07302      0.01657    -4.41  1.1e-05 ***
## Year2010          -0.04986      0.01709    -2.92  0.00354 **
## Year2011          -0.05784      0.01739    -3.33  0.00088 ***
## Year2012          -0.04193      0.01836    -2.28  0.02239 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.00495,    Adjusted R-squared:  0.00413
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 7 observations c(2881,4814,6464,6480,10467,17718,18961)
## are outliers with |weight| = 0 ( < 4.9e-06);
## 1806 weights are ~ = 1. The remaining 18791 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0012 0.8600 0.9500 0.8860 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.85e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 20604"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1304"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 582 604 595 532 580 512 474 435 412 434 450 468 463 500 447
## 2011 2012
## 480 430
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

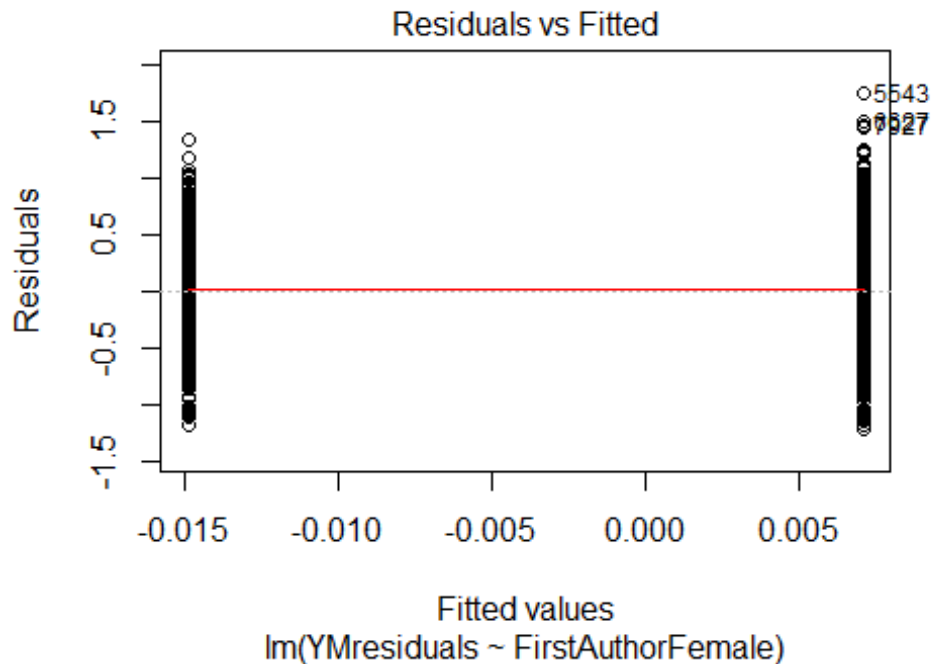
```



```
## 287 327 311 270 311 246 286 223 249 244 298 273 270 308 283
## 2011 2012
## 319 274
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 246 289 278 236 283 219 266 194 218 215 262 240 241 268 243
## 2011 2012
## 269 228
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 3e-04
```

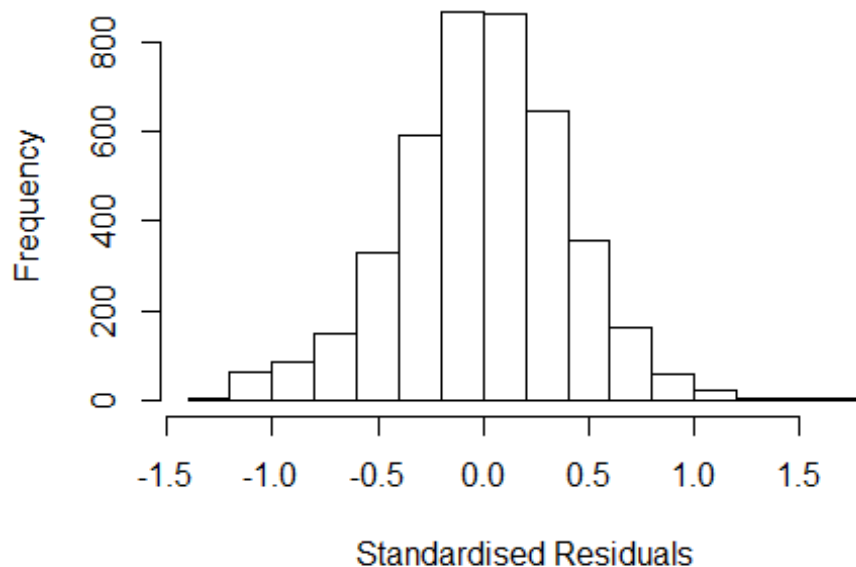


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 26, df = 1, p-value = 4e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.54928050109796"
## [1] "Male first author team size 2018 geometric mean: 3.21956036445795"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8900, p-value = 5e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.99353335337394"
## [1] "Male last author team size 2018 geometric mean: 3.53577789682762"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## LastAuthorFemale  1.032 1      1.016
## UniqueAuthors     1.135 4      1.016
## Year              1.130 16     1.004
```

## Residuals from first and last author and team size



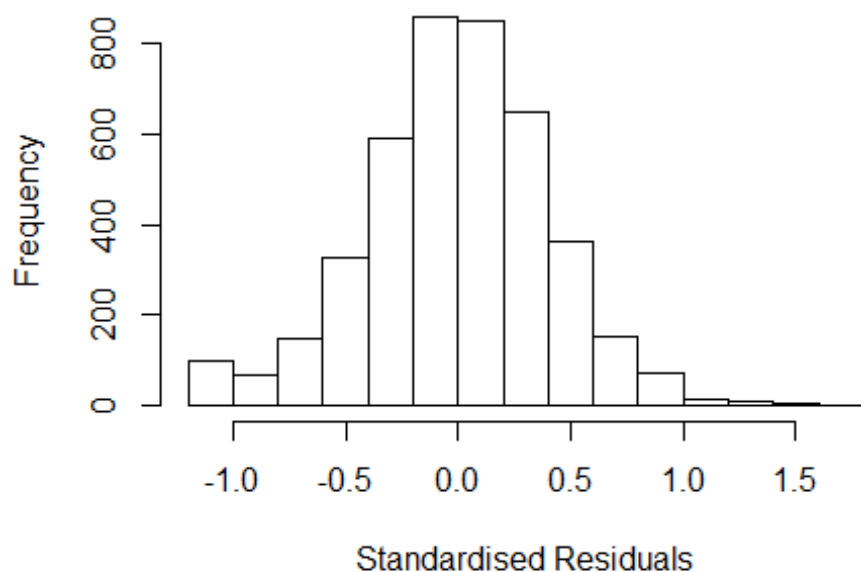
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21043 -0.25127  0.00285  0.25814  1.76739
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01494    0.04197   24.18  < 2e-16 ***
## FirstAuthorFemale1 -0.03801    0.01308   -2.91  0.00367 **
## LastAuthorFemale1 -0.00441    0.01734   -0.25  0.79929
## UniqueAuthors2     0.17272    0.03356    5.15  2.8e-07 ***
## UniqueAuthors3     0.16652    0.03317    5.02  5.4e-07 ***
## UniqueAuthors4     0.17834    0.03337    5.34  9.5e-08 ***
## UniqueAuthors5     0.24347    0.03253    7.48  8.7e-14 ***
## Year1997           0.01600    0.04037    0.40  0.69194
## Year1998          -0.04797    0.03962   -1.21  0.22598
## Year1999          -0.13552    0.03923   -3.45  0.00056 ***
```

```

## Year2000      -0.08925    0.03758   -2.38  0.01758 *
## Year2001      -0.07297    0.03950   -1.85  0.06473 .
## Year2002      -0.09709    0.03750   -2.59  0.00966 **
## Year2003      -0.10019    0.03996   -2.51  0.01221 *
## Year2004      -0.11305    0.03981   -2.84  0.00454 **
## Year2005      -0.08921    0.04089   -2.18  0.02917 *
## Year2006      -0.10684    0.03802   -2.81  0.00498 **
## Year2007      -0.07290    0.03952   -1.84  0.06516 .
## Year2008      -0.10398    0.03950   -2.63  0.00851 **
## Year2009      -0.08988    0.03901   -2.30  0.02127 *
## Year2010      -0.12406    0.03919   -3.17  0.00156 **
## Year2011      -0.13673    0.04029   -3.39  0.00069 ***
## Year2012      -0.11886    0.04115   -2.89  0.00389 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0335, Adjusted R-squared:  0.0284
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 382 weights are ~= 1. The remaining 3813 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000  0.864  0.949  0.896  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.38e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
## trace.lev    mts    compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1 1.009
## LastAuthorFemale 1.018 1 1.009
## Year 1.035 16 1.001

```

## Residuals from first and last author



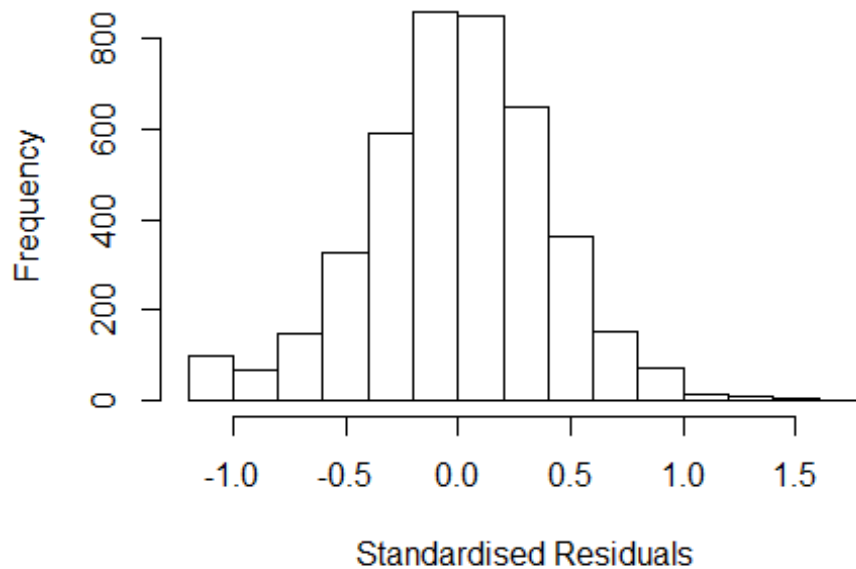
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19667 -0.25089 0.00122 0.25878 1.76065
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.183408 0.030821 38.40 < 2e-16 ***
## FirstAuthorFemale1 -0.020931 0.012962 -1.61 0.10642
## LastAuthorFemale1 -0.000644 0.017339 -0.04 0.97037
## Year1997 0.013262 0.040825 0.32 0.74531
## Year1998 -0.043126 0.040040 -1.08 0.28151
## Year1999 -0.135454 0.039406 -3.44 0.00059 ***
## Year2000 -0.078929 0.037795 -2.09 0.03683 *
## Year2001 -0.074173 0.039560 -1.87 0.06087 .
## Year2002 -0.093406 0.038004 -2.46 0.01402 *
## Year2003 -0.084629 0.039962 -2.12 0.03426 *
## Year2004 -0.102055 0.040164 -2.54 0.01109 *
## Year2005 -0.071600 0.040857 -1.75 0.07977 .
```

```

## Year2006      -0.100600    0.038439   -2.62  0.00890 **
## Year2007      -0.055188    0.039603   -1.39  0.16353
## Year2008      -0.099732    0.039576   -2.52  0.01177 *
## Year2009      -0.080642    0.038915   -2.07  0.03830 *
## Year2010      -0.108574    0.039196   -2.77  0.00563 **
## Year2011      -0.116838    0.040263   -2.90  0.00373 **
## Year2012      -0.104136    0.041655   -2.50  0.01246 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00628
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 356 weights are ~= 1. The remaining 3839 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8650 0.9510 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1      1.009
## Year      1.017 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19660 -0.25077 0.00128 0.25871 1.76074
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1833 0.0306 38.65 < 2e-16 ***
## FirstAuthorFemale1 -0.0210 0.0130 -1.62 0.10623
## Year1997 0.0133 0.0408 0.33 0.74487
## Year1998 -0.0431 0.0400 -1.08 0.28161
## Year1999 -0.1354 0.0394 -3.44 0.00059 ***
## Year2000 -0.0789 0.0378 -2.09 0.03686 *
## Year2001 -0.0742 0.0396 -1.87 0.06090 .
## Year2002 -0.0934 0.0380 -2.46 0.01404 *
## Year2003 -0.0846 0.0399 -2.12 0.03419 *
## Year2004 -0.1021 0.0402 -2.54 0.01110 *
## Year2005 -0.0716 0.0408 -1.75 0.07961 .
## Year2006 -0.1006 0.0384 -2.62 0.00889 **
```

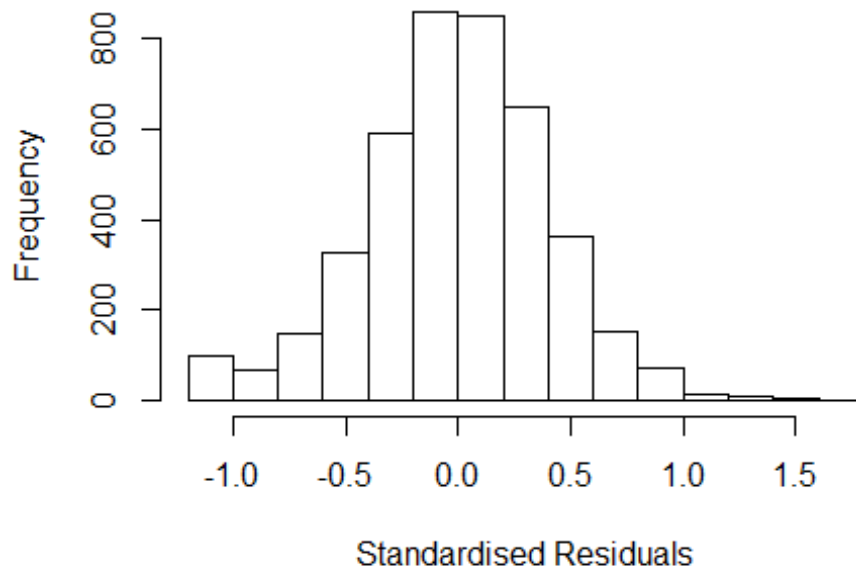
```

## Year2007          -0.0552      0.0396   -1.39   0.16351
## Year2008          -0.0997      0.0396   -2.52   0.01176 *
## Year2009          -0.0806      0.0389   -2.07   0.03832 *
## Year2010          -0.1086      0.0392   -2.77   0.00564 **
## Year2011          -0.1168      0.0403   -2.90   0.00373 **
## Year2012          -0.1041      0.0416   -2.50   0.01244 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00652
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 357 weights are ~= 1. The remaining 3838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8650 0.9510 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.018 1          1.009
## Year            1.018 16          1.001

```



## Residuals from last author



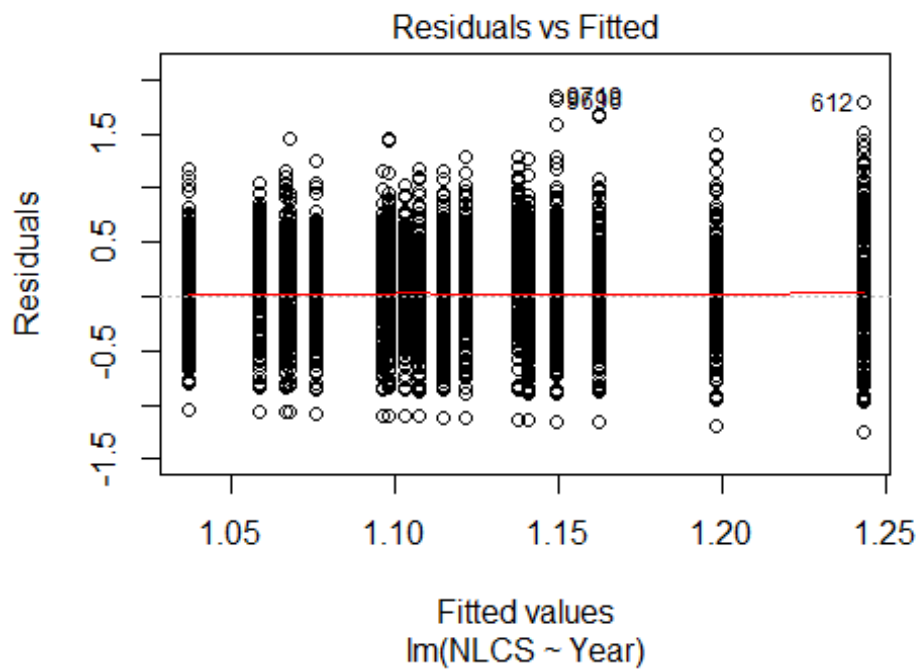
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.191352 -0.252332 0.000853 0.257359 1.768332
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1779 0.0305 38.63 < 2e-16 ***
## LastAuthorFemale1 -0.0021 0.0173 -0.12 0.90369
## Year1997 0.0134 0.0408 0.33 0.74217
## Year1998 -0.0441 0.0400 -1.10 0.27052
## Year1999 -0.1378 0.0393 -3.50 0.00046 ***
## Year2000 -0.0794 0.0377 -2.10 0.03557 *
## Year2001 -0.0757 0.0396 -1.91 0.05596 .
## Year2002 -0.0937 0.0380 -2.47 0.01360 *
## Year2003 -0.0875 0.0399 -2.19 0.02838 *
## Year2004 -0.1043 0.0401 -2.60 0.00939 **
## Year2005 -0.0734 0.0408 -1.80 0.07231 .
## Year2006 -0.1021 0.0384 -2.66 0.00780 **
```

```

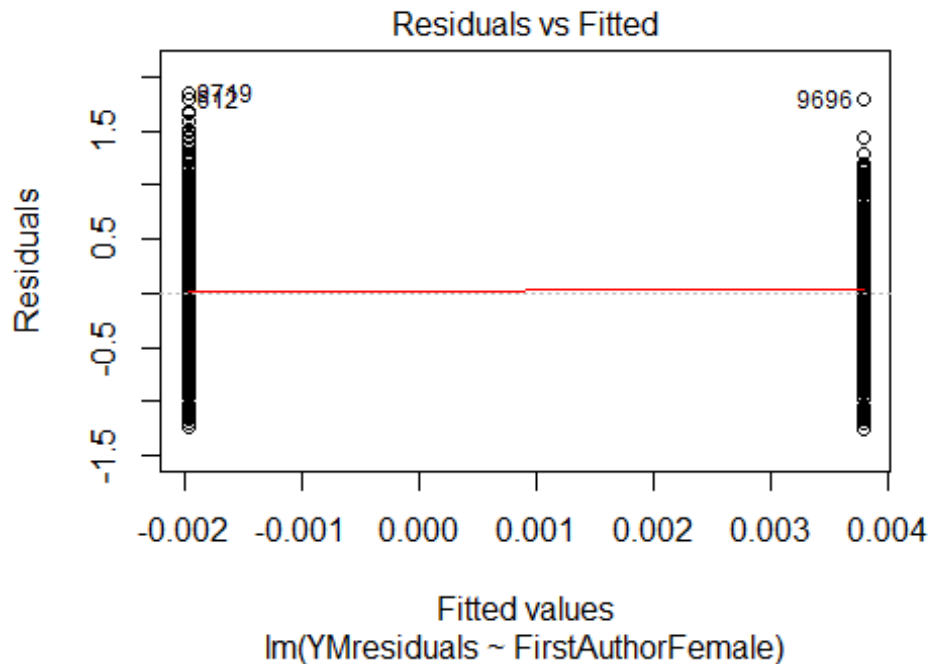
## Year2007          -0.0571      0.0396   -1.44   0.14952
## Year2008          -0.1007      0.0396   -2.54   0.01099 *
## Year2009          -0.0820      0.0389   -2.11   0.03485 *
## Year2010          -0.1111      0.0391   -2.84   0.00449 **
## Year2011          -0.1181      0.0403   -2.93   0.00337 **
## Year2012          -0.1054      0.0416   -2.53   0.01139 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.00994,    Adjusted R-squared:  0.00591
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 369 weights are ~= 1. The remaining 3826 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0003 0.8640 0.9500 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4195"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1305"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 548 563 492 505 544 470 468 386 430 455 463 469 521 511 540
## 2011 2012
## 544 522
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 221 206 195 224 217 173 223 197 244 276 287 277 308 337 338
## 2011 2012

```

```
## 360 338
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 188 174 203 197 152 200 182 207 248 254 247 276 288 294
## 2011 2012
## 325 292
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

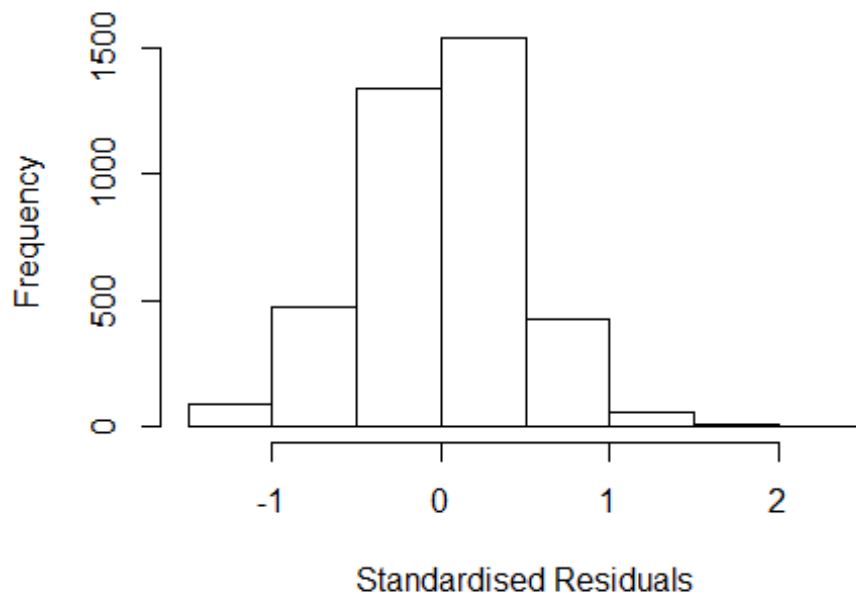


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.8, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 3.84389774586663"
## [1] "Male first author team size 2018 geometric mean: 3.45029045900824"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.27679039666808"
## [1] "Male last author team size 2018 geometric mean: 3.67340976849751"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6400, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## LastAuthorFemale  1.023 1          1.012
## UniqueAuthors    1.151 4          1.018
## Year              1.172 16         1.005
```

## Residuals from first and last author and team size



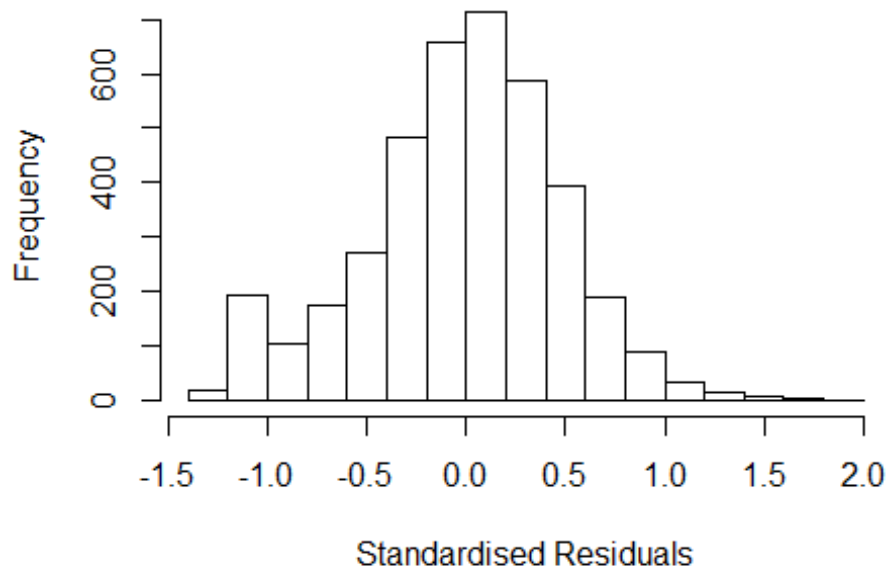
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4413 -0.2843 0.0169 0.2900 2.2541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9272 0.0587 15.80 < 2e-16 ***
## FirstAuthorFemale1 -0.0234 0.0156 -1.51 0.13215
## LastAuthorFemale1 -0.0381 0.0191 -1.99 0.04661 *
## UniqueAuthors2 0.3621 0.0342 10.60 < 2e-16 ***
## UniqueAuthors3 0.4035 0.0323 12.48 < 2e-16 ***
## UniqueAuthors4 0.4443 0.0330 13.46 < 2e-16 ***
## UniqueAuthors5 0.5376 0.0313 17.18 < 2e-16 ***
## Year1997 -0.0823 0.0710 -1.16 0.24641
## Year1998 -0.0833 0.0667 -1.25 0.21215
## Year1999 -0.1597 0.0633 -2.52 0.01176 *
```

```

## Year2000          -0.2136      0.0636   -3.36  0.00080 ***
## Year2001          -0.2191      0.0669   -3.28  0.00106 **
## Year2002          -0.2184      0.0630   -3.47  0.00053 ***
## Year2003          -0.1919      0.0622   -3.09  0.00205 **
## Year2004          -0.2014      0.0609   -3.31  0.00096 ***
## Year2005          -0.1985      0.0616   -3.22  0.00128 **
## Year2006          -0.2146      0.0599   -3.58  0.00035 ***
## Year2007          -0.1843      0.0608   -3.03  0.00246 **
## Year2008          -0.1731      0.0612   -2.83  0.00471 **
## Year2009          -0.1712      0.0598   -2.86  0.00421 **
## Year2010          -0.2329      0.0597   -3.90  9.8e-05 ***
## Year2011          -0.1883      0.0602   -3.13  0.00177 **
## Year2012          -0.1682      0.0596   -2.82  0.00483 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.13,   Adjusted R-squared:  0.125
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 3532 is an outlier with |weight| = 0 ( < 2.5e-05);
## 355 weights are ~= 1. The remaining 3571 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0375 0.8520 0.9500 0.8940 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          2.55e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500          50          2          1          1000          200
## trace.lev mts compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.021 1          1.010
## LastAuthorFemale 1.027 1          1.013
## Year 1.048 16          1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2562 -0.2940 0.0163 0.3079 1.8381
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2390 0.0605 20.48 <2e-16 ***
## FirstAuthorFemale1 0.0172 0.0161 1.07 0.2848
## LastAuthorFemale1 -0.0614 0.0203 -3.03 0.0025 **
## Year1997 -0.0544 0.0783 -0.69 0.4875
## Year1998 -0.0371 0.0728 -0.51 0.6106
## Year1999 -0.1183 0.0686 -1.73 0.0845 .
## Year2000 -0.1824 0.0694 -2.63 0.0086 **
## Year2001 -0.1666 0.0738 -2.26 0.0241 *
## Year2002 -0.1658 0.0687 -2.41 0.0158 *
## Year2003 -0.1040 0.0687 -1.51 0.1303
## Year2004 -0.1239 0.0668 -1.86 0.0636 .
## Year2005 -0.1403 0.0675 -2.08 0.0376 *
```

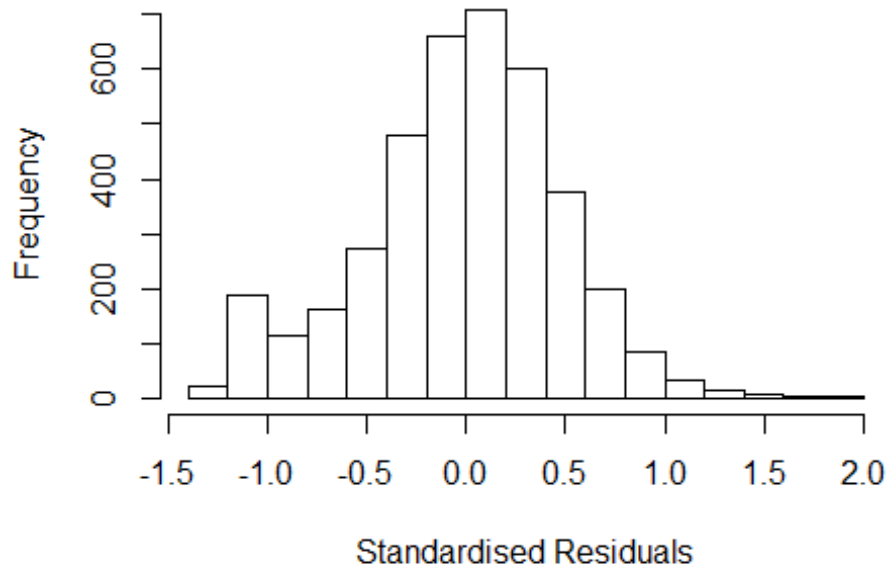
```

## Year2006          -0.1229      0.0656   -1.87   0.0610 .
## Year2007          -0.0788      0.0663   -1.19   0.2348
## Year2008          -0.0784      0.0670   -1.17   0.2421
## Year2009          -0.0908      0.0653   -1.39   0.1647
## Year2010          -0.1353      0.0659   -2.05   0.0402 *
## Year2011          -0.0841      0.0661   -1.27   0.2030
## Year2012          -0.0588      0.0656   -0.90   0.3701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00664
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 363 weights are ~= 1. The remaining 3564 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0546 0.8540 0.9490 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.55e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## Year              1.022 16      1.001

```



## Residuals from first author



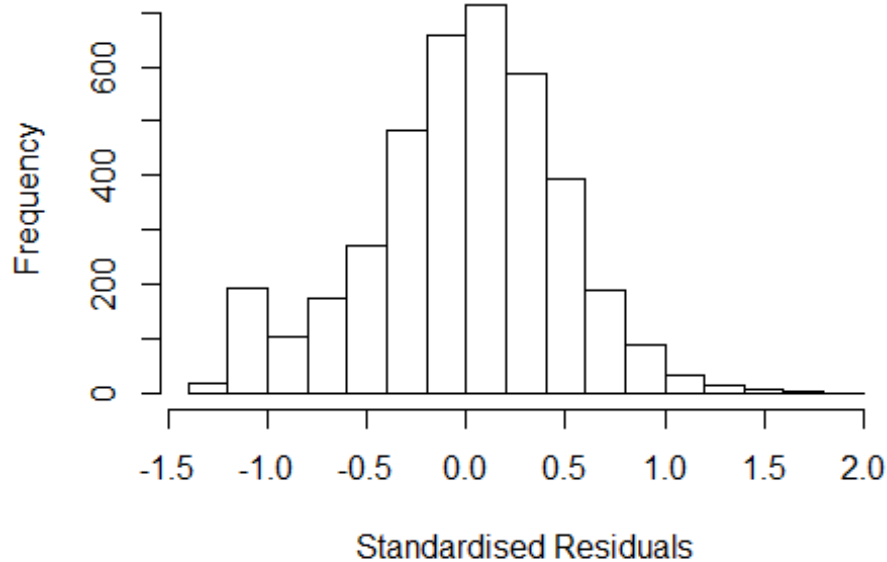
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2452 -0.2956 0.0181 0.3083 1.8499
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2345 0.0605 20.40 <2e-16 ***
## FirstAuthorFemale1 0.0107 0.0162 0.66 0.5106
## Year1997 -0.0519 0.0782 -0.66 0.5073
## Year1998 -0.0402 0.0728 -0.55 0.5808
## Year1999 -0.1216 0.0685 -1.78 0.0759 .
## Year2000 -0.1869 0.0694 -2.69 0.0071 **
## Year2001 -0.1711 0.0740 -2.31 0.0207 *
## Year2002 -0.1692 0.0688 -2.46 0.0139 *
## Year2003 -0.1059 0.0687 -1.54 0.1230
## Year2004 -0.1289 0.0668 -1.93 0.0537 .
## Year2005 -0.1461 0.0675 -2.17 0.0304 *
## Year2006 -0.1262 0.0656 -1.92 0.0545 .
```

```

## Year2007          -0.0856      0.0662   -1.29   0.1960
## Year2008          -0.0856      0.0668   -1.28   0.2000
## Year2009          -0.0945      0.0653   -1.45   0.1480
## Year2010          -0.1405      0.0658   -2.13   0.0330 *
## Year2011          -0.0914      0.0660   -1.39   0.1661
## Year2012          -0.0657      0.0656   -1.00   0.3166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.00888,    Adjusted R-squared:  0.00457
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 367 weights are ~= 1. The remaining 3560 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0501 0.8550 0.9490 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.55e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.026 1          1.013
## Year            1.026 16          1.001

```

## Residuals from last author



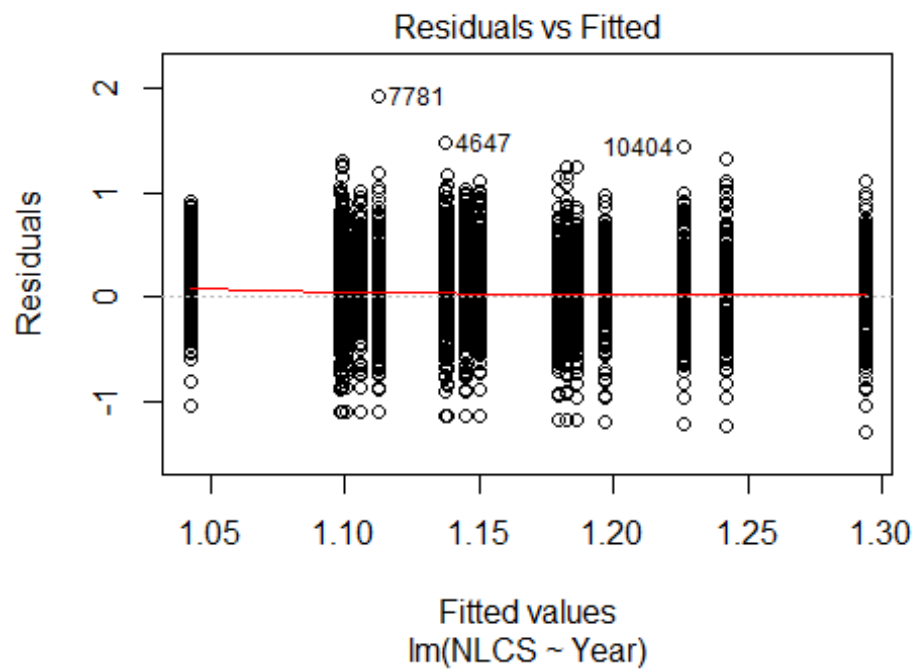
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2447 -0.2935  0.0178  0.3066  1.8325
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2447     0.0600  20.76  <2e-16 ***
## LastAuthorFemale1 -0.0586     0.0204  -2.87   0.0042 **
## Year1997         -0.0545     0.0782  -0.70   0.4861
## Year1998         -0.0388     0.0726  -0.53   0.5936
## Year1999         -0.1196     0.0685  -1.75   0.0807 .
## Year2000         -0.1837     0.0692  -2.65   0.0080 **
## Year2001         -0.1668     0.0739  -2.26   0.0240 *
## Year2002         -0.1661     0.0687  -2.42   0.0157 *
## Year2003         -0.1033     0.0687  -1.50   0.1326
## Year2004         -0.1234     0.0667  -1.85   0.0645 .
## Year2005         -0.1408     0.0674  -2.09   0.0367 *
## Year2006         -0.1225     0.0656  -1.87   0.0620 .
```

```

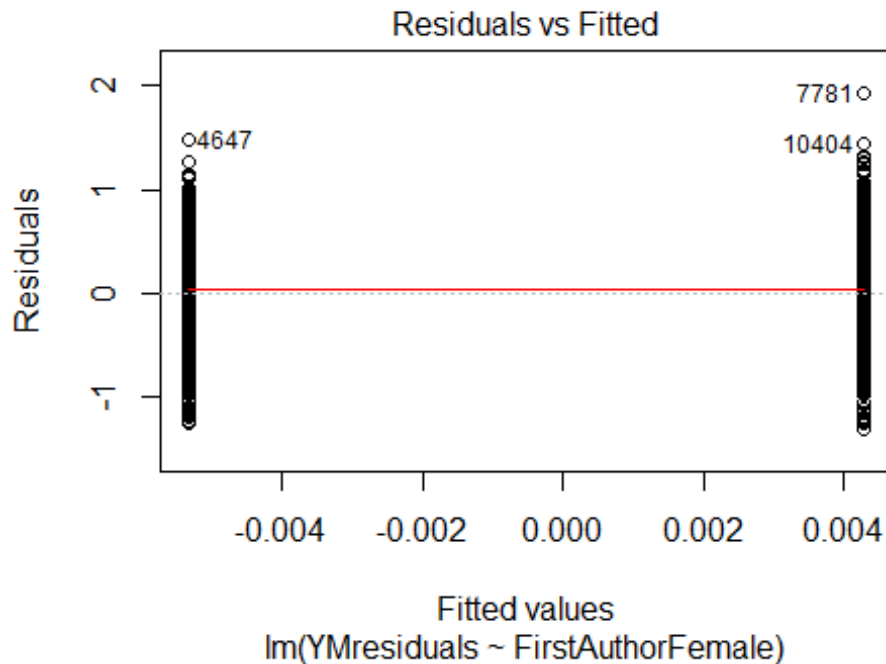
## Year2007          -0.0794      0.0663    -1.20    0.2314
## Year2008          -0.0785      0.0669    -1.17    0.2406
## Year2009          -0.0907      0.0653    -1.39    0.1650
## Year2010          -0.1347      0.0659    -2.04    0.0410 *
## Year2011          -0.0842      0.0660    -1.28    0.2021
## Year2012          -0.0588      0.0655    -0.90    0.3695
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.011, Adjusted R-squared:  0.00671
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 363 weights are ~= 1. The remaining 3564 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0567 0.8550 0.9480 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.55e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3927"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1306"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 655 638 520 576 632 591 578 519 536 548 602 642 614 630 563
## 2011 2012
## 586 552
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 199 173 222 258 196 274 239 250 274 290 363 332 337 324
## 2011 2012

```

```
## 337 305
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 180 182 150 196 231 174 223 200 214 230 261 319 293 293 275
## 2011 2012
## 299 270
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 80, df = 16, p-value = 2e-10
```

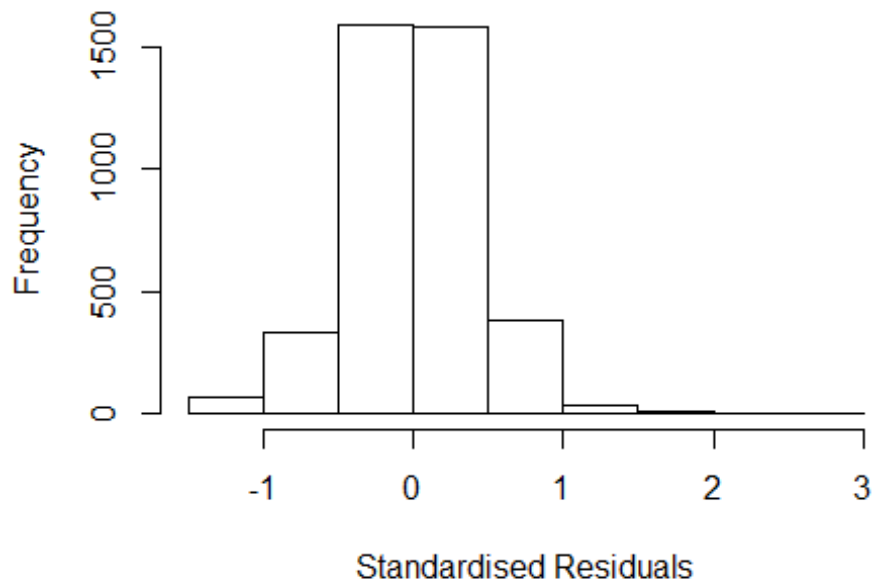


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 6.13085624990743"
## [1] "Male first author team size 2018 geometric mean: 6.33339927985307"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8800, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.72161758876357"
## [1] "Male last author team size 2018 geometric mean: 6.05260489316808"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1      1.020
## LastAuthorFemale  1.035 1      1.017
## UniqueAuthors    1.097 4      1.012
## Year              1.143 16      1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7781 33646413672 3.041 2006      1306      3      2.554
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4094 -0.2664  0.0019  0.2616  2.5545
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.66969    0.04867   13.76 < 2e-16 ***
## FirstAuthorFemale1 -0.04196    0.01317   -3.19  0.00145 **
## LastAuthorFemale1 -0.03481    0.01459   -2.39  0.01708 *
## UniqueAuthors2     0.58141    0.04154   14.00 < 2e-16 ***
## UniqueAuthors3     0.60942    0.03979   15.32 < 2e-16 ***
## UniqueAuthors4     0.60776    0.04037   15.05 < 2e-16 ***
## UniqueAuthors5     0.78166    0.03553   22.00 < 2e-16 ***
## Year1997         -0.00117    0.04807   -0.02  0.98064
## Year1998         -0.04354    0.04962   -0.88  0.38033
## Year1999         -0.14810    0.04404   -3.36  0.00078 ***
```

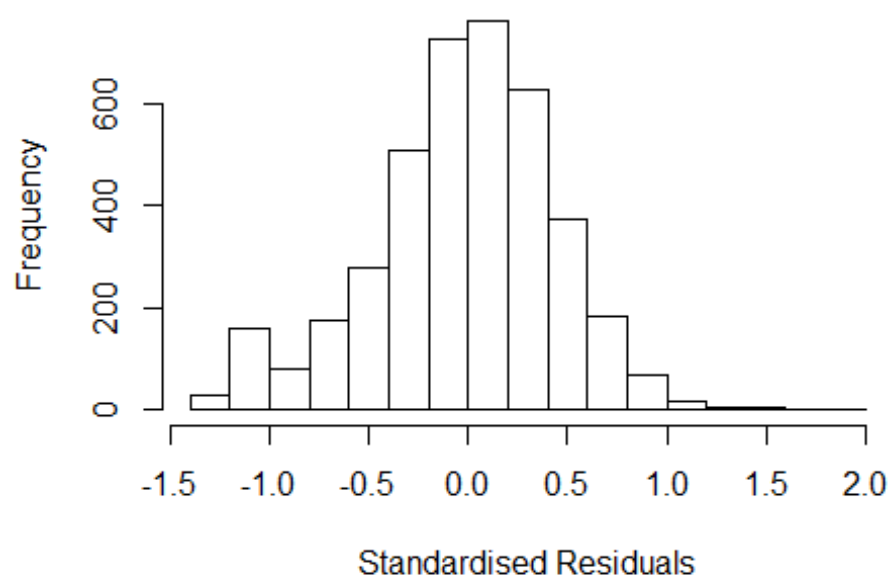
```

## Year2000      -0.20527    0.04331   -4.74  2.2e-06 ***
## Year2001      -0.11455    0.04888   -2.34  0.01916 *
## Year2002      -0.12356    0.04462   -2.77  0.00565 **
## Year2003      -0.19634    0.04222   -4.65  3.4e-06 ***
## Year2004      -0.20637    0.04296   -4.80  1.6e-06 ***
## Year2005      -0.14565    0.04182   -3.48  0.00050 ***
## Year2006      -0.18315    0.04292   -4.27  2.0e-05 ***
## Year2007      -0.19434    0.04282   -4.54  5.8e-06 ***
## Year2008      -0.14466    0.04334   -3.34  0.00085 ***
## Year2009      -0.09362    0.04062   -2.30  0.02124 *
## Year2010      -0.16199    0.04406   -3.68  0.00024 ***
## Year2011      -0.12584    0.04222   -2.98  0.00290 **
## Year2012      -0.19623    0.04278   -4.59  4.6e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.237, Adjusted R-squared:  0.232
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(584,2116)
## are outliers with |weight| <= 5.5e-07 ( < 2.5e-05);
## 325 weights are ~ = 1. The remaining 3663 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0769 0.8660 0.9490 0.8980 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.035 1 1.018
## LastAuthorFemale 1.056 1 1.028
## Year 1.087 16 1.003

```



## Residuals from first and last author

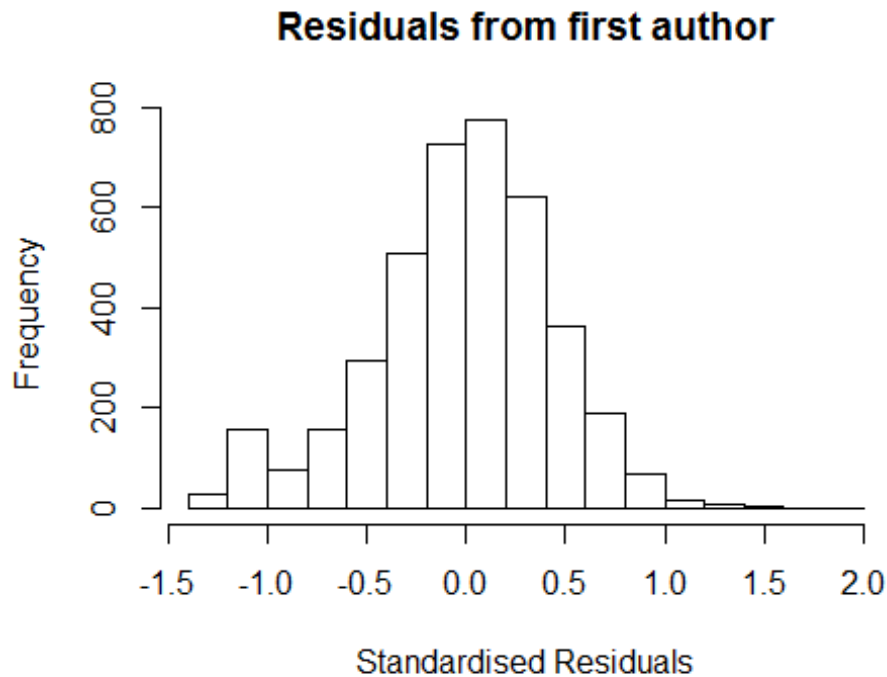


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.317 -0.271  0.012  0.285  1.895
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.30081    0.03953   32.91 < 2e-16 ***
## FirstAuthorFemale1 -0.00842    0.01420   -0.59  0.55315
## LastAuthorFemale1 -0.04702    0.01653   -2.84  0.00447 **
## Year1997         0.01669    0.05110    0.33  0.74403
## Year1998        -0.04334    0.05120   -0.85  0.39737
## Year1999        -0.10831    0.04775   -2.27  0.02337 *
## Year2000        -0.20185    0.05396   -3.74  0.00019 ***
## Year2001        -0.13301    0.05416   -2.46  0.01410 *
## Year2002        -0.10578    0.04921   -2.15  0.03165 *
## Year2003        -0.16151    0.04714   -3.43  0.00062 ***
## Year2004        -0.16545    0.04800   -3.45  0.00057 ***
## Year2005        -0.08179    0.04690   -1.74  0.08125 .
```

```

## Year2006          -0.15524    0.04692   -3.31  0.00095 ***
## Year2007          -0.14936    0.04874   -3.06  0.00219 **
## Year2008          -0.12557    0.04862   -2.58  0.00984 **
## Year2009          -0.04033    0.04493   -0.90  0.36945
## Year2010          -0.11578    0.04870   -2.38  0.01749 *
## Year2011          -0.08038    0.04636   -1.73  0.08301 .
## Year2012          -0.12999    0.04746   -2.74  0.00619 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0183, Adjusted R-squared:  0.0139
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 312 weights are ~= 1. The remaining 3678 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8640 0.9510 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1      1.018
## Year      1.036 16      1.001

```



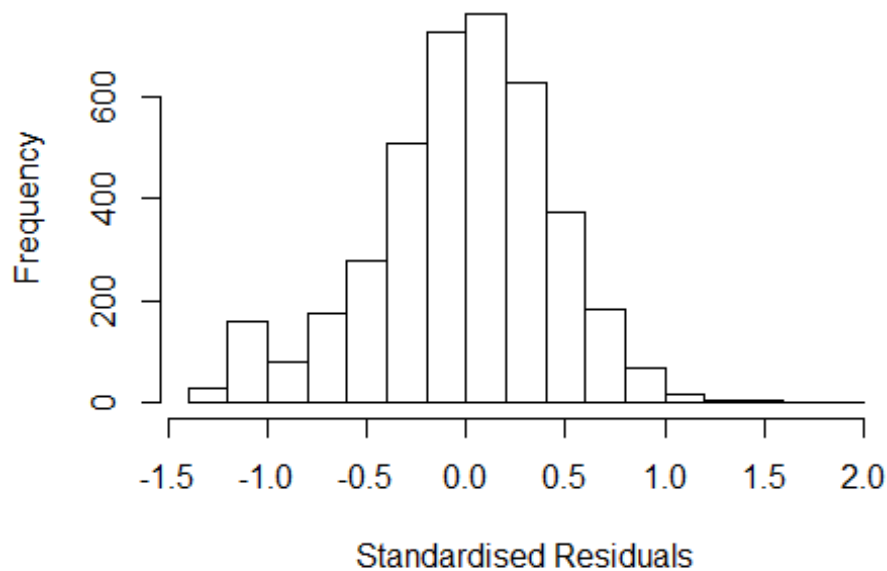
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3104 -0.2738  0.0108  0.2842  1.9034
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2927     0.0394  32.77 < 2e-16 ***
## FirstAuthorFemale1 -0.0123     0.0144  -0.86  0.39148
## Year1997          0.0177     0.0511   0.35  0.72903
## Year1998         -0.0426     0.0512  -0.83  0.40508
## Year1999         -0.1084     0.0477  -2.27  0.02299 *
## Year2000         -0.2012     0.0541  -3.72  0.00020 ***
## Year2001         -0.1347     0.0542  -2.49  0.01295 *
## Year2002         -0.1044     0.0491  -2.13  0.03361 *
## Year2003         -0.1614     0.0472  -3.42  0.00064 ***
## Year2004         -0.1681     0.0481  -3.49  0.00048 ***
## Year2005         -0.0858     0.0468  -1.83  0.06687 .
## Year2006         -0.1551     0.0469  -3.31  0.00096 ***
```

```

## Year2007          -0.1531      0.0487   -3.14  0.00168 **
## Year2008          -0.1263      0.0487   -2.59  0.00959 **
## Year2009          -0.0443      0.0449   -0.99  0.32327
## Year2010          -0.1209      0.0486   -2.49  0.01292 *
## Year2011          -0.0850      0.0464   -1.83  0.06678 .
## Year2012          -0.1349      0.0473   -2.85  0.00437 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.0117
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 319 weights are ~= 1. The remaining 3671 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0011 0.8640 0.9500 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.055 1          1.027
## Year            1.055 16          1.002

```

## Residuals from last author



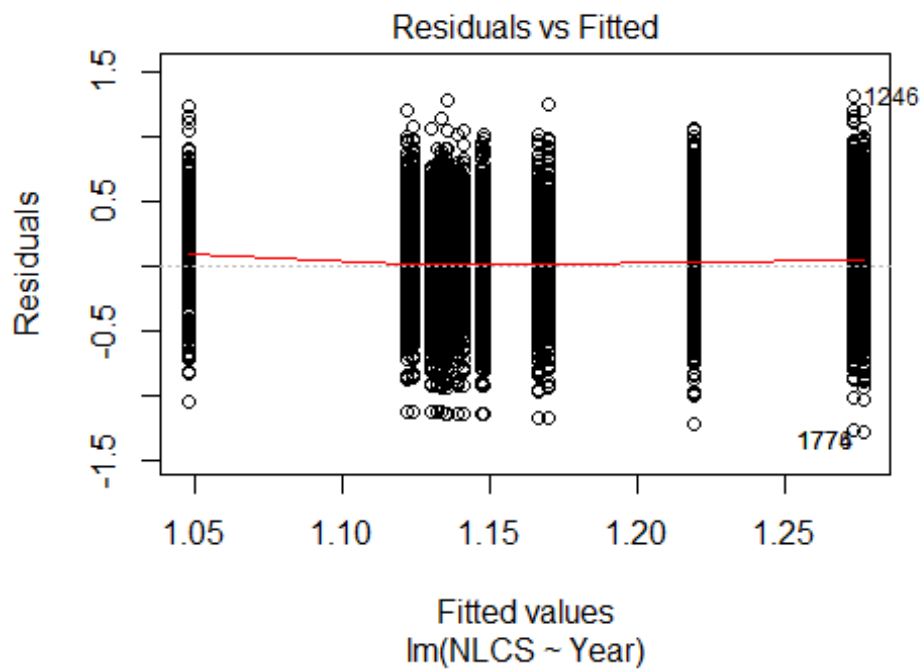
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3146 -0.2744  0.0144  0.2842  1.8990
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2975     0.0391  33.22 < 2e-16 ***
## LastAuthorFemale1 -0.0480     0.0167  -2.87  0.00415 **
## Year1997          0.0172     0.0511   0.34  0.73723
## Year1998         -0.0427     0.0512  -0.83  0.40396
## Year1999         -0.1085     0.0478  -2.27  0.02323 *
## Year2000         -0.2018     0.0540  -3.74  0.00019 ***
## Year2001         -0.1334     0.0542  -2.46  0.01392 *
## Year2002         -0.1062     0.0492  -2.16  0.03104 *
## Year2003         -0.1618     0.0472  -3.43  0.00061 ***
## Year2004         -0.1653     0.0480  -3.44  0.00058 ***
## Year2005         -0.0824     0.0469  -1.76  0.07895 .
## Year2006         -0.1554     0.0470  -3.31  0.00094 ***
```

```

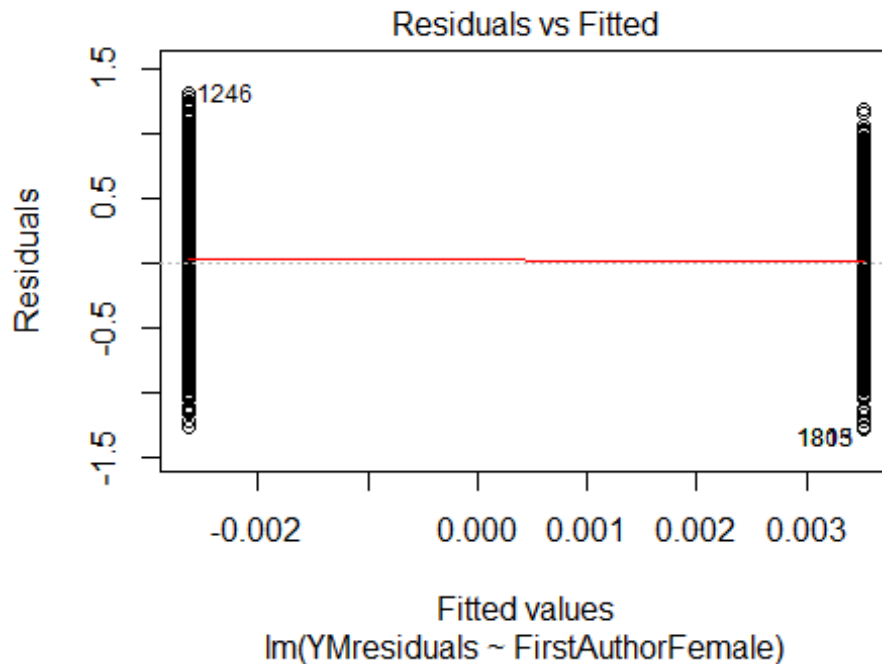
## Year2007          -0.1499      0.0487   -3.08  0.00212 **
## Year2008          -0.1257      0.0486   -2.58  0.00979 **
## Year2009          -0.0410      0.0450   -0.91  0.36129
## Year2010          -0.1161      0.0487   -2.38  0.01722 *
## Year2011          -0.0810      0.0464   -1.75  0.08065 .
## Year2012          -0.1307      0.0475   -2.75  0.00592 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.014
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 312 weights are ~= 1. The remaining 3678 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0015 0.8640 0.9510 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3990"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1307"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1312 1165 1185 1099 1268 1137 1214 970 955 1021 917 992 1009 960 967
## 2011 2012
## 962 832
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 777 688 752 712 628 515 868 676 711 672 648 733 718 723 726
## 2011 2012

```

```
## 746 622
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 680 625 674 637 573 463 790 584 615 592 584 642 643 652 645
## 2011 2012
## 650 541
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 410, df = 16, p-value <2e-16
```



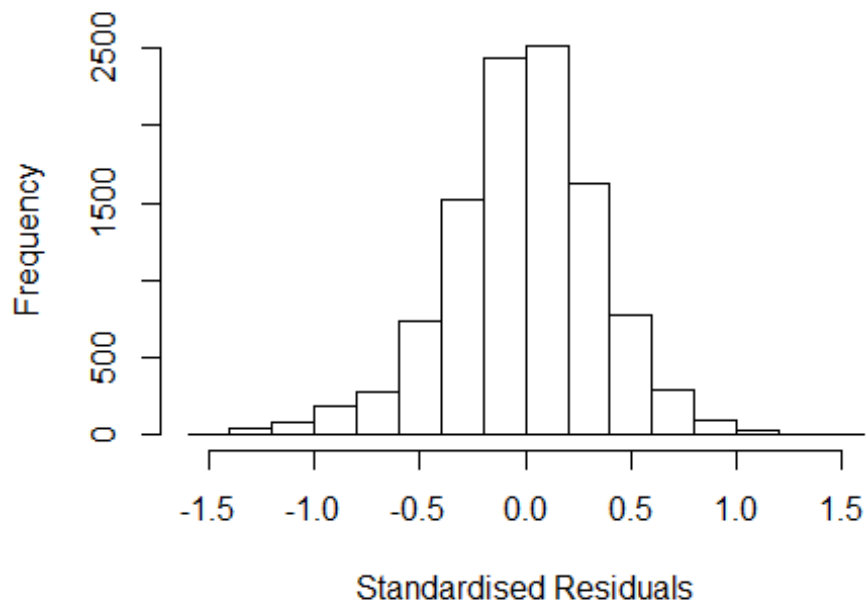
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 54, df = 1, p-value = 2e-13
```



```
## [1] "Female first author team size 2018 geometric mean: 5.15468122130963"
## [1] "Male first author team size 2018 geometric mean: 4.12733549793715"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.95143002433093"
## [1] "Male last author team size 2018 geometric mean: 4.46162350164699"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.021 1          1.010
## LastAuthorFemale  1.020 1          1.010
## UniqueAuthors    1.060 4          1.007
## Year             1.080 16          1.002
```



## Residuals from first and last author and team size



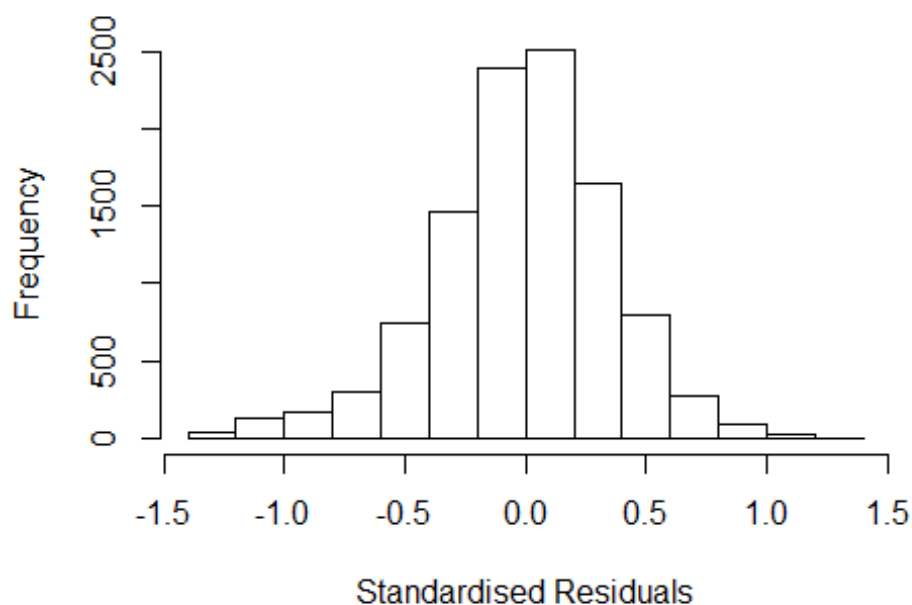
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40316 -0.21966 0.00316 0.21435 1.49329
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08371 0.03158 34.31 < 2e-16 ***
## FirstAuthorFemale1 -0.00359 0.00680 -0.53 0.5971
## LastAuthorFemale1 0.00479 0.00837 0.57 0.5670
## UniqueAuthors2 0.18163 0.02811 6.46 1.1e-10 ***
## UniqueAuthors3 0.20921 0.02781 7.52 5.8e-14 ***
## UniqueAuthors4 0.22358 0.02783 8.03 1.0e-15 ***
## UniqueAuthors5 0.30250 0.02716 11.14 < 2e-16 ***
## Year1997 0.01694 0.02505 0.68 0.4990
## Year1998 -0.07211 0.02261 -3.19 0.0014 **
## Year1999 -0.16246 0.02172 -7.48 8.1e-14 ***
```

```

## Year2000      -0.17938      0.02479      -7.24      5.0e-13 ***
## Year2001      -0.15624      0.02350      -6.65      3.1e-11 ***
## Year2002      -0.14130      0.02069      -6.83      9.0e-12 ***
## Year2003      -0.17000      0.02159      -7.87      3.8e-15 ***
## Year2004      -0.18709      0.02102      -8.90      < 2e-16 ***
## Year2005      -0.18497      0.02083      -8.88      < 2e-16 ***
## Year2006      -0.16958      0.02163      -7.84      4.9e-15 ***
## Year2007      -0.18744      0.02142      -8.75      < 2e-16 ***
## Year2008      -0.17058      0.02181      -7.82      5.8e-15 ***
## Year2009      -0.18055      0.02113      -8.55      < 2e-16 ***
## Year2010      -0.17104      0.02205      -7.76      9.6e-15 ***
## Year2011      -0.17751      0.02258      -7.86      4.1e-15 ***
## Year2012      -0.18503      0.02362      -7.83      5.1e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.326
## Multiple R-squared:  0.0663, Adjusted R-squared:  0.0643
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 873 weights are ~= 1. The remaining 9717 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0017 0.8630 0.9510 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.44e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.016 1          1.008
## LastAuthorFemale 1.017 1          1.008
## Year      1.026 16          1.001

```

## Residuals from first and last author



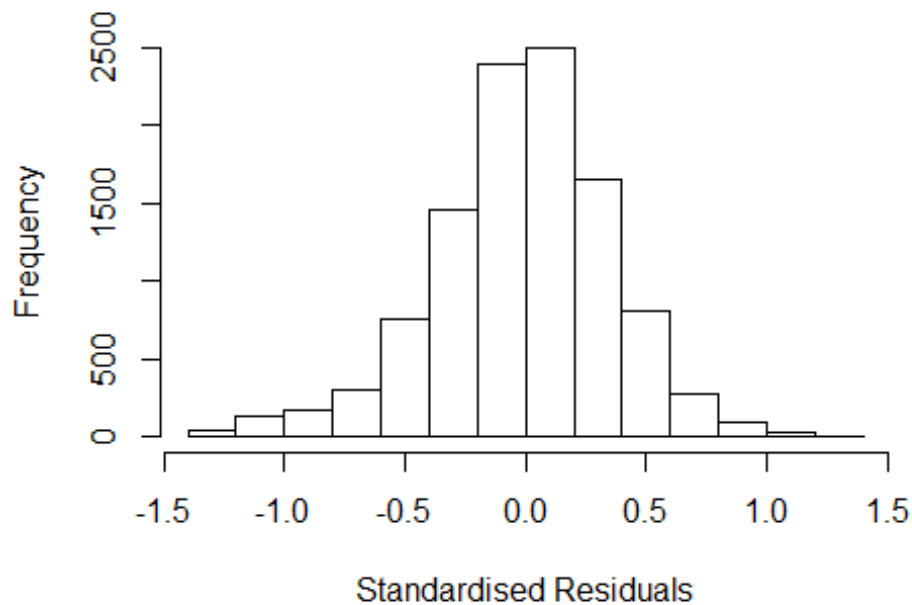
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32629 -0.22149  0.00322  0.21837  1.27200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.30500    0.01787   73.03  < 2e-16 ***
## FirstAuthorFemale1 0.00385    0.00687    0.56  0.5753
## LastAuthorFemale1 0.00160    0.00846    0.19  0.8502
## Year1997         0.01585    0.02497    0.63  0.5258
## Year1998        -0.06889    0.02275   -3.03  0.0025 **
## Year1999        -0.15783    0.02225   -7.09  1.4e-12 ***
## Year2000        -0.18273    0.02553   -7.16  8.8e-13 ***
## Year2001        -0.14365    0.02392   -6.01  2.0e-09 ***
## Year2002        -0.11944    0.02120   -5.63  1.8e-08 ***
## Year2003        -0.15153    0.02192   -6.91  5.0e-12 ***
## Year2004        -0.16758    0.02139   -7.83  5.2e-15 ***
## Year2005        -0.15908    0.02115   -7.52  5.8e-14 ***
```

```

## Year2006          -0.14949      0.02189      -6.83      9.0e-12 ***
## Year2007          -0.16600      0.02166      -7.67      1.9e-14 ***
## Year2008          -0.14817      0.02213      -6.70      2.3e-11 ***
## Year2009          -0.15946      0.02151      -7.41      1.3e-13 ***
## Year2010          -0.14376      0.02238      -6.42      1.4e-10 ***
## Year2011          -0.15412      0.02319      -6.64      3.2e-11 ***
## Year2012          -0.15440      0.02405      -6.42      1.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.0241
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 931 weights are ~= 1. The remaining 9659 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0679 0.8600 0.9500 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.44e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.013 1      1.006
## Year      1.013 16      1.000

```

## Residuals from first author



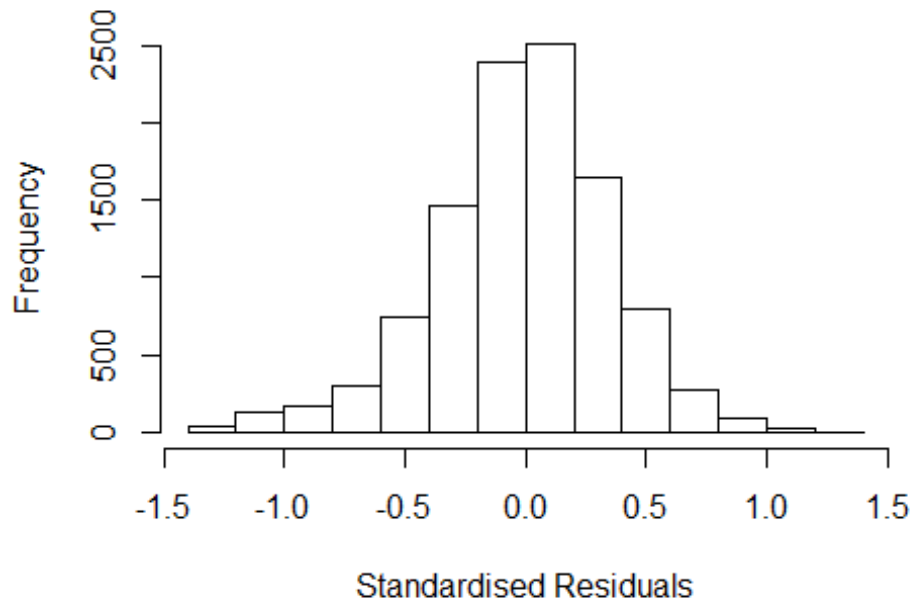
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32499 -0.22126  0.00308  0.21830  1.27174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.30526    0.01778   73.41  < 2e-16 ***
## FirstAuthorFemale1 0.00396    0.00686    0.58  0.5637
## Year1997        0.01576    0.02496    0.63  0.5278
## Year1998       -0.06893    0.02274   -3.03  0.0024 **
## Year1999       -0.15788    0.02225   -7.10  1.4e-12 ***
## Year2000       -0.18268    0.02553   -7.15  9.0e-13 ***
## Year2001       -0.14363    0.02392   -6.00  2.0e-09 ***
## Year2002       -0.11946    0.02119   -5.64  1.8e-08 ***
## Year2003       -0.15153    0.02192   -6.91  5.0e-12 ***
## Year2004       -0.16757    0.02140   -7.83  5.3e-15 ***
## Year2005       -0.15902    0.02115   -7.52  6.0e-14 ***
## Year2006       -0.14949    0.02189   -6.83  9.0e-12 ***
```

```

## Year2007          -0.16597    0.02166   -7.66  2.0e-14 ***
## Year2008          -0.14812    0.02213   -6.69  2.3e-11 ***
## Year2009          -0.15942    0.02150   -7.41  1.3e-13 ***
## Year2010          -0.14370    0.02238   -6.42  1.4e-10 ***
## Year2011          -0.15405    0.02320   -6.64  3.3e-11 ***
## Year2012          -0.15430    0.02404   -6.42  1.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.0242
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 936 weights are ~= 1. The remaining 9654 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0685 0.8600 0.9500 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.44e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.014 1          1.007
## Year            1.014 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32440 -0.22096 0.00294 0.21829 1.27072
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.30628 0.01771 73.74 < 2e-16 ***
## LastAuthorFemale1 0.00201 0.00845 0.24 0.8121
## Year1997 0.01611 0.02496 0.65 0.5187
## Year1998 -0.06881 0.02275 -3.02 0.0025 **
## Year1999 -0.15745 0.02224 -7.08 1.5e-12 ***
## Year2000 -0.18264 0.02553 -7.15 9.0e-13 ***
## Year2001 -0.14334 0.02392 -5.99 2.1e-09 ***
## Year2002 -0.11922 0.02120 -5.62 1.9e-08 ***
## Year2003 -0.15120 0.02192 -6.90 5.6e-12 ***
## Year2004 -0.16706 0.02137 -7.82 5.9e-15 ***
## Year2005 -0.15857 0.02112 -7.51 6.5e-14 ***
## Year2006 -0.14903 0.02187 -6.81 1.0e-11 ***
```

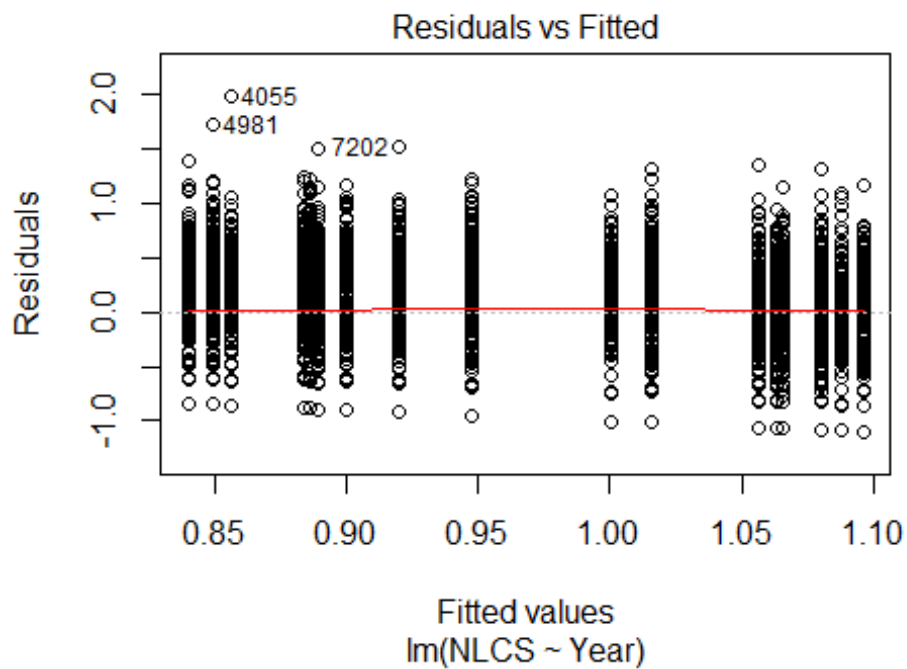
```

## Year2007          -0.16566      0.02165      -7.65   2.1e-14 ***
## Year2008          -0.14781      0.02212      -6.68   2.5e-11 ***
## Year2009          -0.15890      0.02148      -7.40   1.5e-13 ***
## Year2010          -0.14326      0.02236      -6.41   1.6e-10 ***
## Year2011          -0.15361      0.02316      -6.63   3.5e-11 ***
## Year2012          -0.15387      0.02403      -6.40   1.6e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.329
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.0242
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 938 weights are ~= 1. The remaining 9652 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0689 0.8600 0.9500 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.44e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 10590"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1308"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 450 392 386 393 416 422 353 426 410 374 405 403 453 402 385
## 2011 2012
## 420 347
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 180 195 185 167 174 195 272 279 259 285 288 333 306 283
## 2011 2012

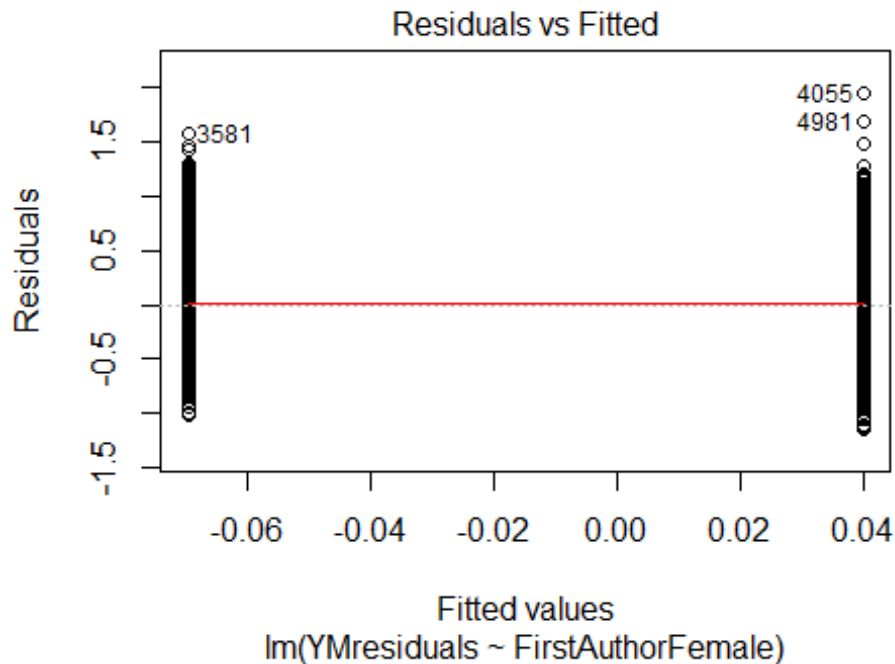
```



```
## 304 251
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 177 161 179 166 143 155 164 232 249 220 262 259 287 278 246
## 2011 2012
## 269 216
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 68, df = 16, p-value = 3e-08
```

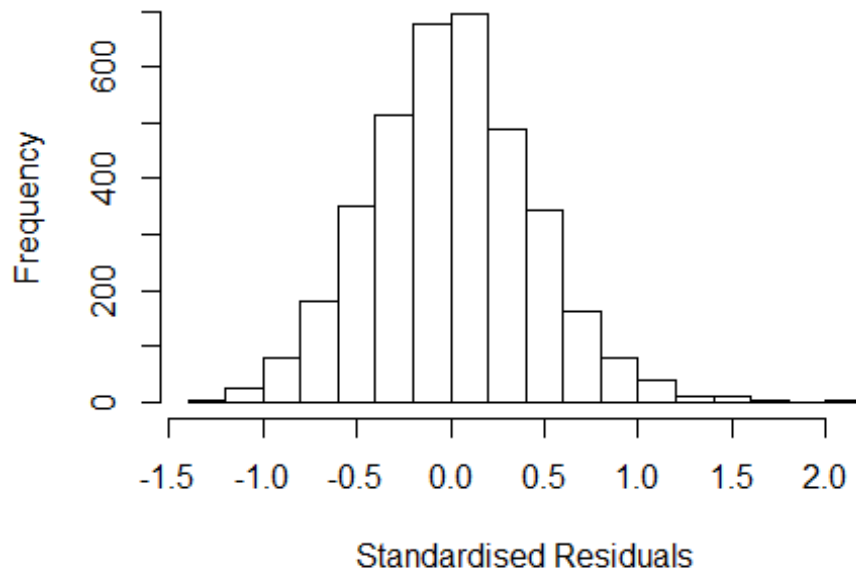


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 20, df = 1, p-value = 6e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 5.23226845583699"
## [1] "Male first author team size 2018 geometric mean: 4.4882293156223"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.59202972844122"
## [1] "Male last author team size 2018 geometric mean: 4.42152191608304"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.095 1          1.046
## LastAuthorFemale  1.151 1          1.073
## UniqueAuthors    1.407 4          1.044
## Year              1.311 16         1.008
```

## Residuals from first and last author and team size



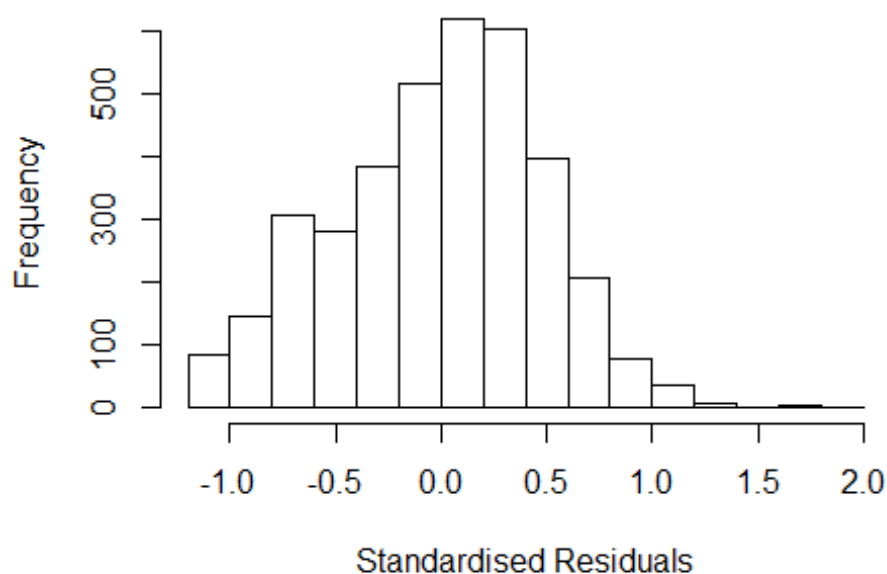
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30400 -0.28082  0.00327  0.28216  2.12411
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.656471   0.046000   14.27 < 2e-16 ***
## FirstAuthorFemale1 -0.070173   0.015815   -4.44 9.4e-06 ***
## LastAuthorFemale1 -0.121755   0.018694   -6.51 8.4e-11 ***
## UniqueAuthors2     0.246435   0.033782    7.29 3.7e-13 ***
## UniqueAuthors3     0.436964   0.034052   12.83 < 2e-16 ***
## UniqueAuthors4     0.547112   0.033882   16.15 < 2e-16 ***
## UniqueAuthors5     0.673980   0.027158   24.82 < 2e-16 ***
## Year1997          -0.059147   0.054980   -1.08 0.28209
## Year1998          -0.006744   0.052381   -0.13 0.89756
## Year1999          -0.026453   0.049779   -0.53 0.59517
```

```

## Year2000      -0.061823    0.053579    -1.15    0.24863
## Year2001      0.002515    0.049493     0.05    0.95948
## Year2002      0.000449    0.048319     0.01    0.99258
## Year2003     -0.147896    0.046214    -3.20    0.00139 **
## Year2004     -0.148903    0.045262    -3.29    0.00101 **
## Year2005     -0.184006    0.046682    -3.94    8.2e-05 ***
## Year2006     -0.209581    0.044735    -4.68    2.9e-06 ***
## Year2007     -0.164139    0.046476    -3.53    0.00042 ***
## Year2008     -0.124237    0.044878    -2.77    0.00566 **
## Year2009     -0.174078    0.046190    -3.77    0.00017 ***
## Year2010     -0.177274    0.047477    -3.73    0.00019 ***
## Year2011     -0.202392    0.046088    -4.39    1.2e-05 ***
## Year2012     -0.107057    0.049415    -2.17    0.03034 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.292, Adjusted R-squared:  0.287
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1480,1924) are outliers with |weight| = 0 ( < 2.7e-05);
## 323 weights are ~ = 1. The remaining 3338 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0312 0.8550 0.9470 0.8940 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.73e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1 1.012
## LastAuthorFemale 1.021 1 1.010
## Year 1.035 16 1.001

```

## Residuals from first and last author



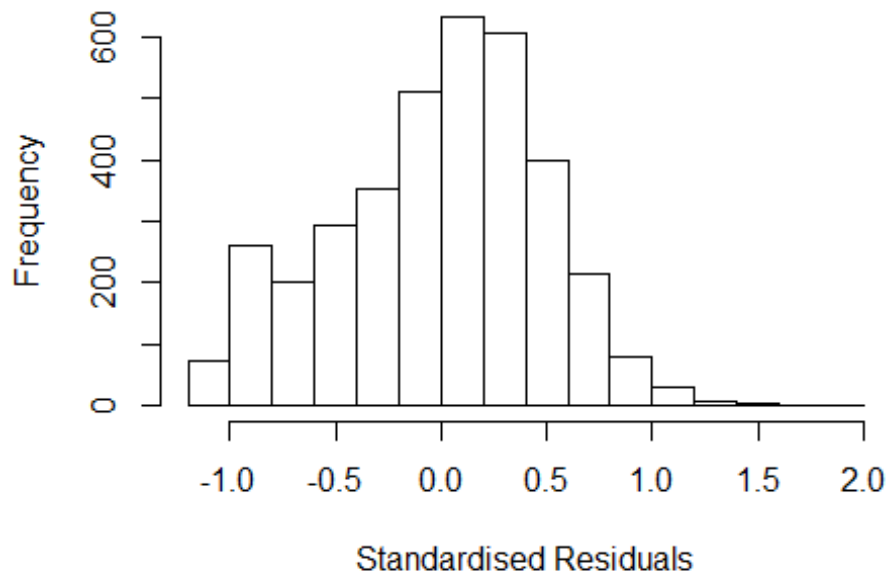
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1814 -0.3438 0.0424 0.3296 1.8897
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15546 0.03834 30.14 < 2e-16 ***
## FirstAuthorFemale1 -0.08569 0.01783 -4.81 1.6e-06 ***
## LastAuthorFemale1 -0.20282 0.02235 -9.07 < 2e-16 ***
## Year1997 -0.05506 0.05901 -0.93 0.35089
## Year1998 -0.02997 0.05393 -0.56 0.57841
## Year1999 -0.02649 0.04956 -0.53 0.59297
## Year2000 -0.03513 0.05314 -0.66 0.50857
## Year2001 0.00888 0.05073 0.18 0.86105
## Year2002 0.02594 0.05002 0.52 0.60403
## Year2003 -0.18956 0.04861 -3.90 9.8e-05 ***
## Year2004 -0.20320 0.04872 -4.17 3.1e-05 ***
## Year2005 -0.23758 0.05221 -4.55 5.5e-06 ***
```

```

## Year2006          -0.23169      0.04974      -4.66   3.3e-06 ***
## Year2007          -0.16068      0.05220      -3.08   0.00210 **
## Year2008          -0.09337      0.04864      -1.92   0.05498 .
## Year2009          -0.13582      0.04944      -2.75   0.00604 **
## Year2010          -0.13101      0.05207      -2.52   0.01191 *
## Year2011          -0.17068      0.04953      -3.45   0.00058 ***
## Year2012          -0.02785      0.05131      -0.54   0.58731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0748, Adjusted R-squared:  0.0702
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 282 weights are ~= 1. The remaining 3381 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.074  0.853   0.947   0.903   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year              1.024 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1551 -0.3610 0.0483 0.3291 1.9317
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12328 0.03777 29.74 < 2e-16 ***
## FirstAuthorFemale1 -0.11737 0.01881 -6.24 4.9e-10 ***
## Year1997 -0.05336 0.05803 -0.92 0.35784
## Year1998 -0.02170 0.05325 -0.41 0.68370
## Year1999 -0.02227 0.04883 -0.46 0.64833
## Year2000 -0.02076 0.05160 -0.40 0.68743
## Year2001 0.00879 0.04969 0.18 0.85952
## Year2002 0.03184 0.04888 0.65 0.51487
## Year2003 -0.19201 0.04948 -3.88 0.00011 ***
## Year2004 -0.21294 0.04914 -4.33 1.5e-05 ***
## Year2005 -0.24686 0.05414 -4.56 5.3e-06 ***
## Year2006 -0.23211 0.05093 -4.56 5.3e-06 ***
```

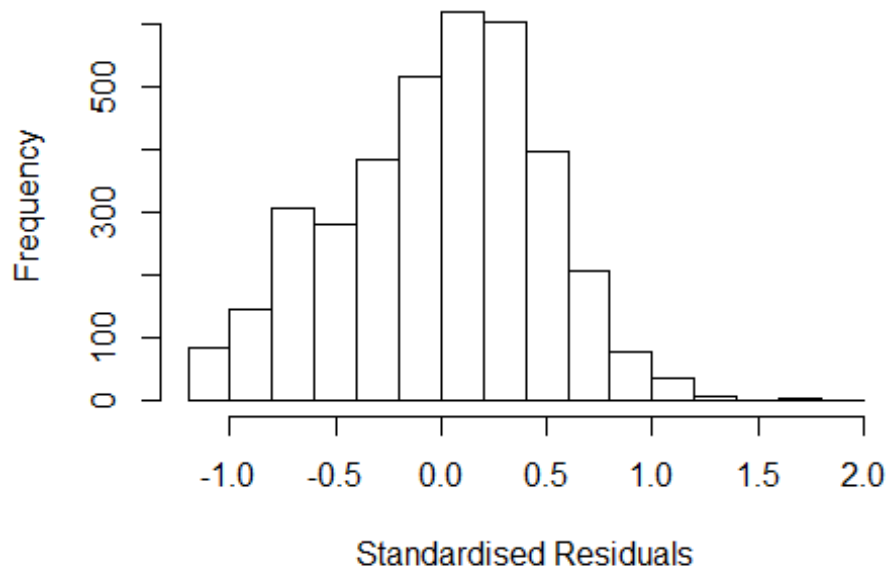
```

## Year2007          -0.17165      0.05310      -3.23  0.00124 **
## Year2008          -0.10324      0.04902      -2.11  0.03524 *
## Year2009          -0.15237      0.05027      -3.03  0.00245 **
## Year2010          -0.13263      0.05258      -2.52  0.01169 *
## Year2011          -0.17554      0.04930      -3.56  0.00037 ***
## Year2012          -0.02720      0.05095      -0.53  0.59346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0448, Adjusted R-squared:  0.0404
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 271 weights are ~= 1. The remaining 3392 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.057  0.855  0.945  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1      1.006
## Year      1.012 16      1.000

```



## Residuals from last author



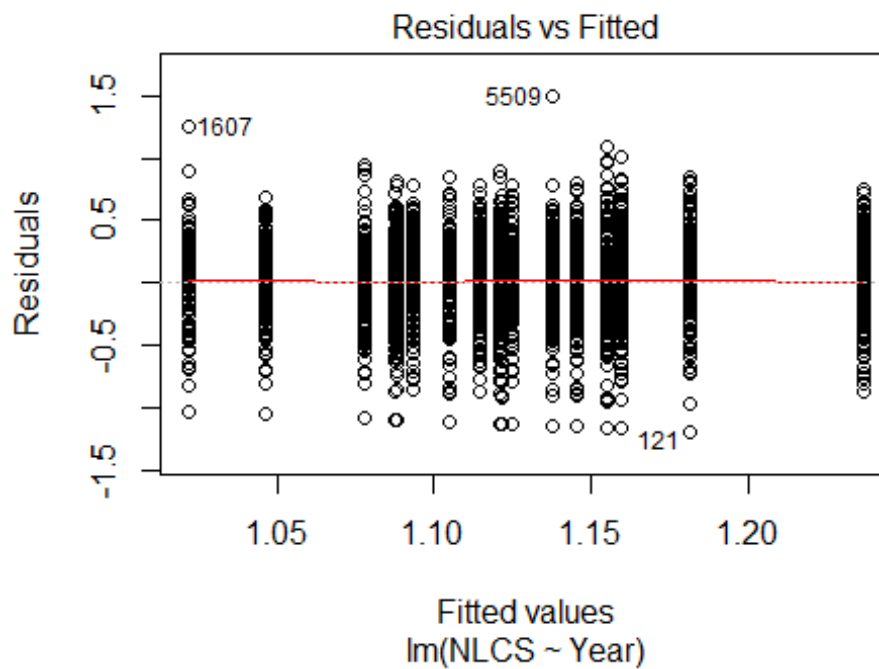
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1561 -0.3392  0.0405  0.3318  1.9160
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13594    0.03762   30.20 < 2e-16 ***
## LastAuthorFemale1 -0.22123    0.02305  -9.60 < 2e-16 ***
## Year1997        -0.05874    0.05861  -1.00  0.31629
## Year1998        -0.03758    0.05338  -0.70  0.48152
## Year1999        -0.02950    0.04914  -0.60  0.54830
## Year2000        -0.04393    0.05294  -0.83  0.40665
## Year2001         0.00572    0.05036   0.11  0.90952
## Year2002         0.02014    0.04932   0.41  0.68295
## Year2003        -0.19702    0.04884  -4.03  5.6e-05 ***
## Year2004        -0.20996    0.04866  -4.31  1.6e-05 ***
## Year2005        -0.25024    0.05232  -4.78  1.8e-06 ***
## Year2006        -0.24045    0.04981  -4.83  1.4e-06 ***
```

```

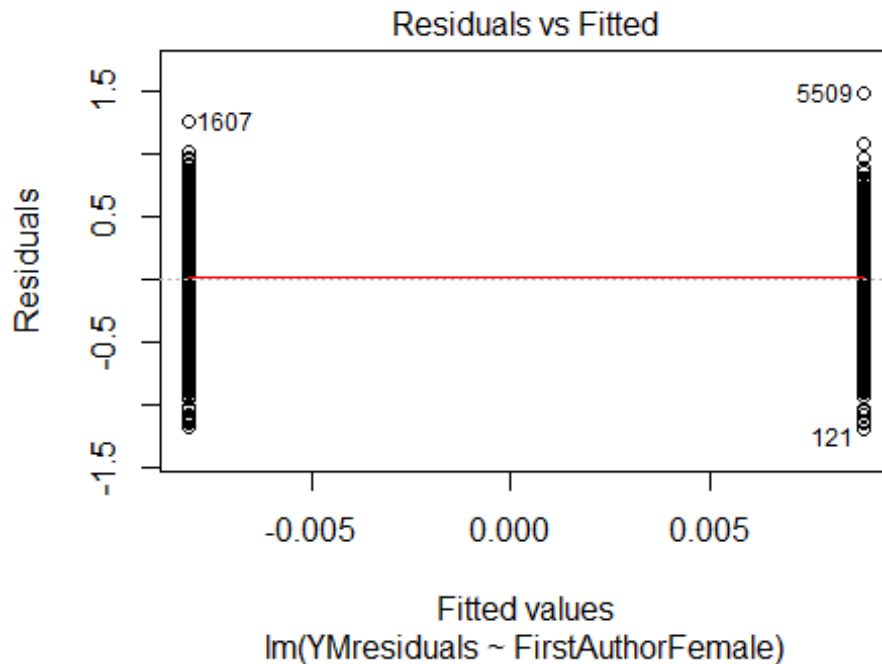
## Year2007          -0.16647      0.05200      -3.20   0.00138 **
## Year2008          -0.10054      0.04871      -2.06   0.03908 *
## Year2009          -0.14329      0.04949      -2.90   0.00381 **
## Year2010          -0.14749      0.05192      -2.84   0.00452 **
## Year2011          -0.17519      0.04936      -3.55   0.00039 ***
## Year2012          -0.03798      0.05124      -0.74   0.45856
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.473
## Multiple R-squared:  0.0674, Adjusted R-squared:  0.0631
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 280 weights are ~= 1. The remaining 3383 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0632 0.8510 0.9450 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3663"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1309"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 313 318 275 269 321 273 234 234 222 256 238 285 257 251 291
## 2011 2012
## 252 251
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 160 158 147 138 138 129 158 169 159 185 166 198 171 178 207
## 2011 2012

```

```
## 176 187
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 139 145 135 127 128 122 144 150 139 159 143 178 146 164 189
## 2011 2012
## 160 168
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 54, df = 16, p-value = 5e-06
```

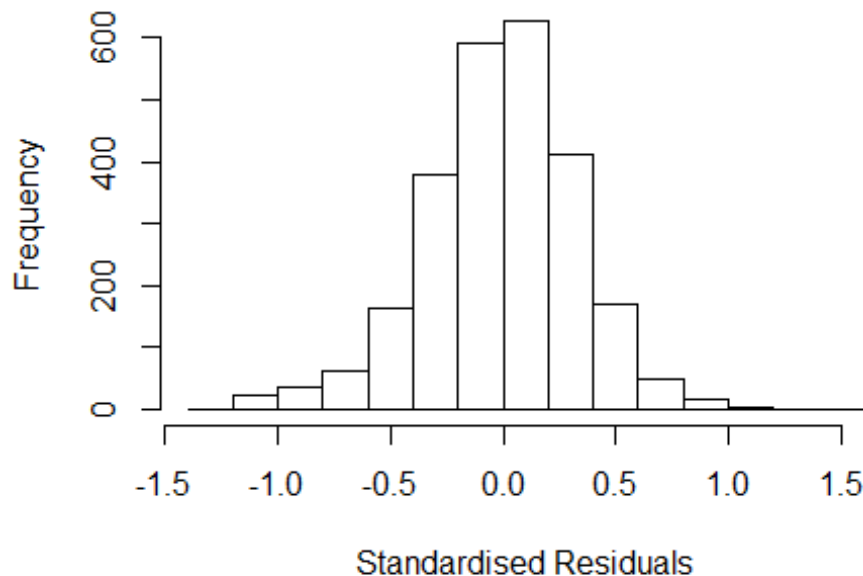


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 3e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.33669707429545"
## [1] "Male first author team size 2018 geometric mean: 3.4660142439667"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.07165218326658"
## [1] "Male last author team size 2018 geometric mean: 3.82932693773622"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1          1.025
## LastAuthorFemale  1.045 1          1.022
## UniqueAuthors    1.224 4          1.026
## Year             1.263 16          1.007
```

## Residuals from first and last author and team size



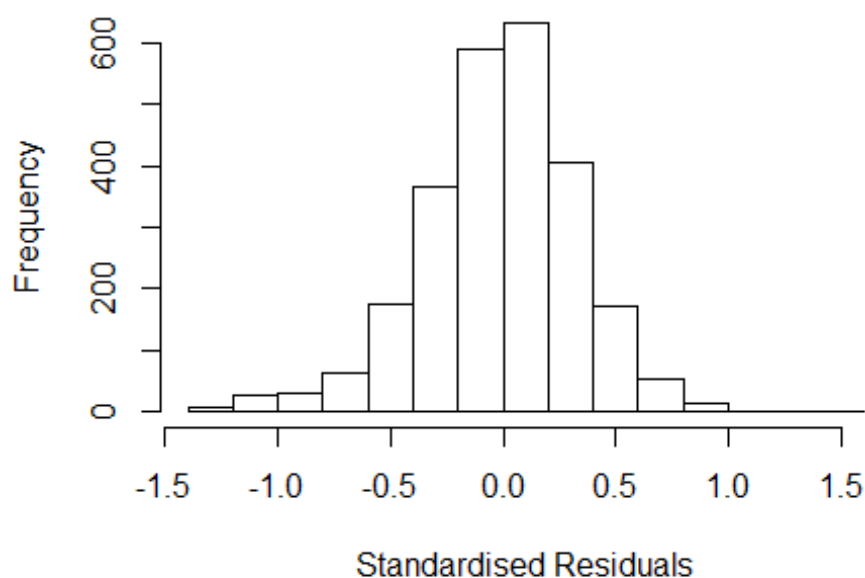
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23856 -0.20930  0.00572  0.20719  1.48905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12154    0.04074   27.53 < 2e-16 ***
## FirstAuthorFemale1  0.00341    0.01330    0.26  0.79752
## LastAuthorFemale1  0.00359    0.01444    0.25  0.80396
## UniqueAuthors2     0.11361    0.03116    3.65  0.00027 ***
## UniqueAuthors3     0.10604    0.03117    3.40  0.00068 ***
## UniqueAuthors4     0.12826    0.03215    3.99  6.8e-05 ***
## UniqueAuthors5     0.16748    0.03037    5.51  3.9e-08 ***
## Year1997           0.02615    0.04229    0.62  0.53638
## Year1998          -0.08606    0.05007   -1.72  0.08580 .
## Year1999          -0.10841    0.04025   -2.69  0.00712 **
```

```

## Year2000      -0.19437      0.04482      -4.34      1.5e-05 ***
## Year2001      -0.13882      0.03964      -3.50      0.00047 ***
## Year2002      -0.08540      0.04065      -2.10      0.03575 *
## Year2003      -0.17501      0.03851      -4.54      5.8e-06 ***
## Year2004      -0.13825      0.04050      -3.41      0.00065 ***
## Year2005      -0.12609      0.04106      -3.07      0.00216 **
## Year2006      -0.09283      0.04018      -2.31      0.02095 *
## Year2007      -0.15647      0.03874      -4.04      5.5e-05 ***
## Year2008      -0.15422      0.04002      -3.85      0.00012 ***
## Year2009      -0.18887      0.03930      -4.81      1.6e-06 ***
## Year2010      -0.11315      0.04064      -2.78      0.00540 **
## Year2011      -0.09961      0.04105      -2.43      0.01530 *
## Year2012      -0.10199      0.04035      -2.53      0.01154 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.314
## Multiple R-squared:  0.0462, Adjusted R-squared:  0.0379
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2289 is an outlier with |weight| = 0 ( < 3.9e-05);
## 236 weights are ~= 1. The remaining 2299 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0258 0.8680 0.9500 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev mts compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1      1.015
## LastAuthorFemale 1.046 1      1.023
## Year      1.069 16      1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24254 -0.21107  0.00281  0.20674  1.46941
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22653    0.03128   39.21 < 2e-16 ***
## FirstAuthorFemale1  0.01272    0.01327    0.96  0.33812
## LastAuthorFemale1  0.00329    0.01456    0.23  0.82123
## Year1997         0.02397    0.04254    0.56  0.57310
## Year1998        -0.08426    0.04922   -1.71  0.08701 .
## Year1999        -0.10927    0.04072   -2.68  0.00733 **
## Year2000        -0.19218    0.04511   -4.26  2.1e-05 ***
## Year2001        -0.13522    0.03956   -3.42  0.00064 ***
## Year2002        -0.08342    0.04116   -2.03  0.04280 *
## Year2003        -0.16384    0.03876   -4.23  2.4e-05 ***
## Year2004        -0.12491    0.04092   -3.05  0.00230 **
## Year2005        -0.11355    0.04189   -2.71  0.00676 **
```

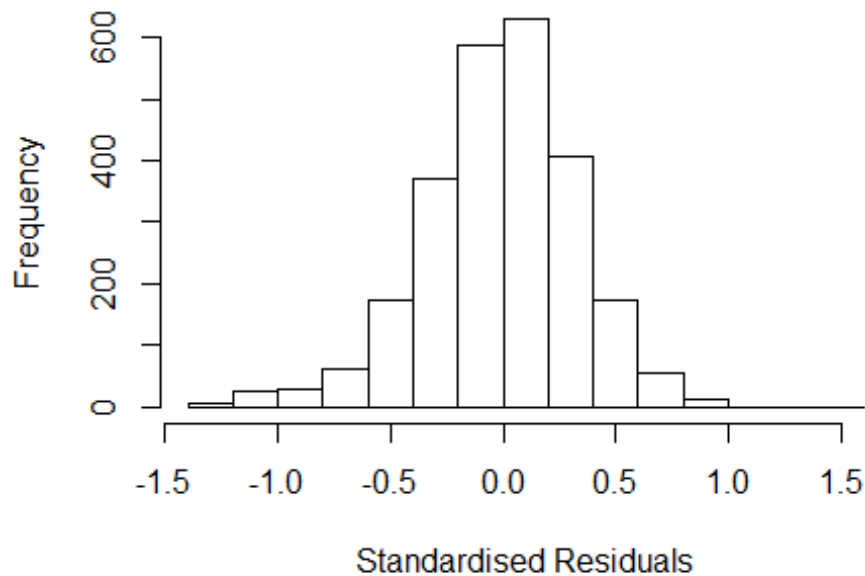
```

## Year2006          -0.08040    0.03965   -2.03  0.04272 *
## Year2007          -0.13820    0.03878   -3.56  0.00037 ***
## Year2008          -0.13697    0.03973   -3.45  0.00058 ***
## Year2009          -0.16795    0.03930   -4.27  2.0e-05 ***
## Year2010          -0.08607    0.04049   -2.13  0.03363 *
## Year2011          -0.08067    0.04078   -1.98  0.04803 *
## Year2012          -0.07757    0.04004   -1.94  0.05284 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.314
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.0198
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 2289 is an outlier with |weight| <= 1.2e-07 ( < 3.9e-05);
## 230 weights are ~= 1. The remaining 2305 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0813 0.8660 0.9520 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1          1.013
## Year              1.027 16          1.001

```



## Residuals from first author



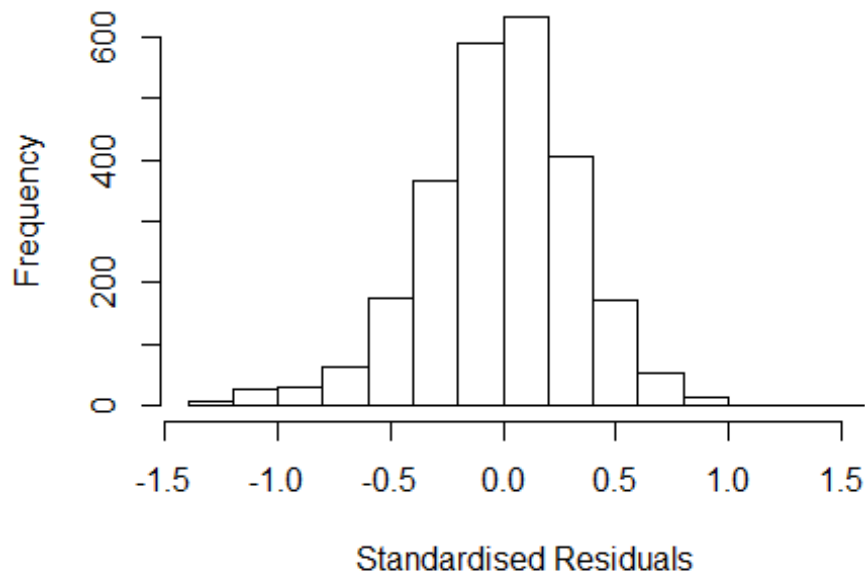
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24035 -0.21143 0.00335 0.20754 1.46806
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2273 0.0312 39.39 < 2e-16 ***
## FirstAuthorFemale1 0.0130 0.0133 0.98 0.32544
## Year1997 0.0237 0.0425 0.56 0.57630
## Year1998 -0.0841 0.0492 -1.71 0.08771 .
## Year1999 -0.1094 0.0407 -2.69 0.00726 **
## Year2000 -0.1921 0.0451 -4.26 2.1e-05 ***
## Year2001 -0.1351 0.0395 -3.42 0.00064 ***
## Year2002 -0.0835 0.0412 -2.03 0.04264 *
## Year2003 -0.1638 0.0388 -4.23 2.5e-05 ***
## Year2004 -0.1250 0.0409 -3.05 0.00228 **
## Year2005 -0.1133 0.0418 -2.71 0.00680 **
## Year2006 -0.0803 0.0397 -2.03 0.04292 *
```

```

## Year2007          -0.1382      0.0388   -3.56  0.00037 ***
## Year2008          -0.1370      0.0397   -3.45  0.00058 ***
## Year2009          -0.1678      0.0393   -4.27  2.0e-05 ***
## Year2010          -0.0860      0.0405   -2.13  0.03360 *
## Year2011          -0.0804      0.0407   -1.97  0.04853 *
## Year2012          -0.0773      0.0400   -1.93  0.05324 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.314
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.0201
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2289 is an outlier with |weight| <= 2.3e-06 ( < 3.9e-05);
## 233 weights are ~= 1. The remaining 2302 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0825 0.8660 0.9510 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.042 1          1.021
## Year            1.042 16          1.001

```

## Residuals from last author



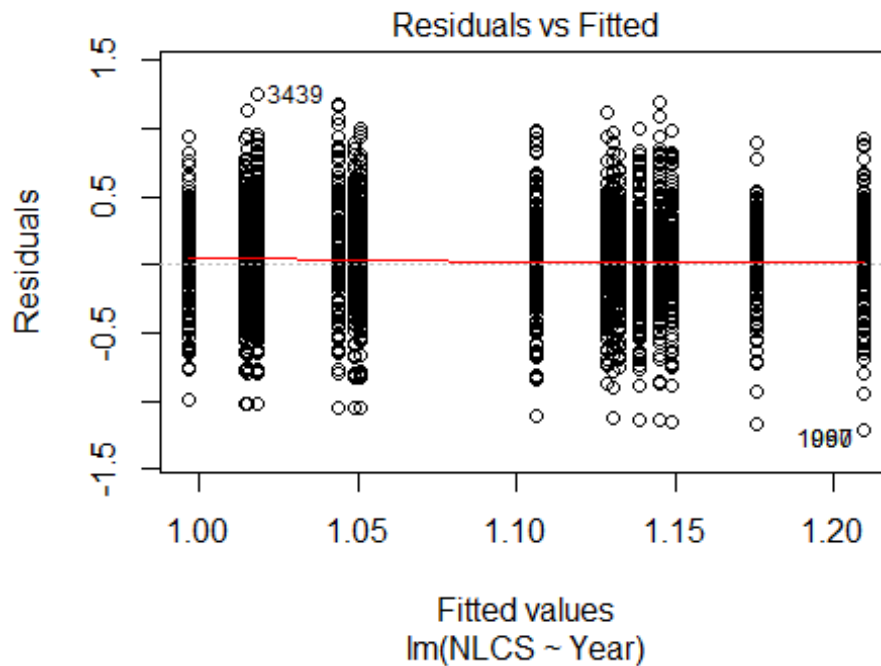
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23702 -0.21312  0.00186  0.20434  1.47535
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23219    0.03097   39.78 < 2e-16 ***
## LastAuthorFemale1 0.00483    0.01453    0.33  0.73963
## Year1997        0.02350    0.04258    0.55  0.58097
## Year1998       -0.08407    0.04925   -1.71  0.08797 .
## Year1999       -0.10887    0.04065   -2.68  0.00745 **
## Year2000       -0.19260    0.04522   -4.26  2.1e-05 ***
## Year2001       -0.13606    0.03961   -3.44  0.00060 ***
## Year2002       -0.08332    0.04120   -2.02  0.04324 *
## Year2003       -0.16264    0.03875   -4.20  2.8e-05 ***
## Year2004       -0.12461    0.04096   -3.04  0.00237 **
## Year2005       -0.11293    0.04201   -2.69  0.00723 **
## Year2006       -0.08095    0.03961   -2.04  0.04107 *
```

```

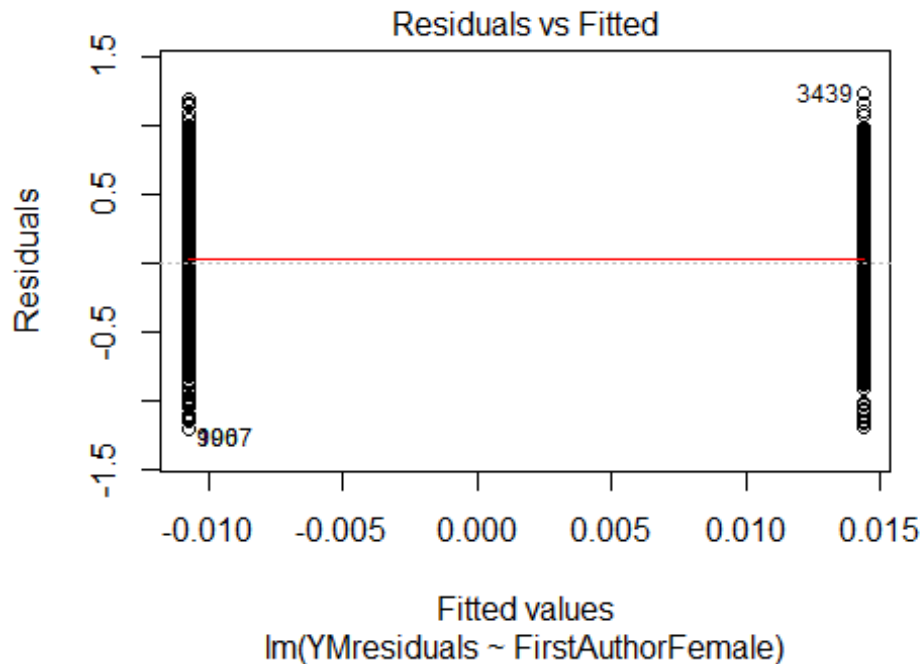
## Year2007          -0.13865      0.03880      -3.57   0.00036 ***
## Year2008          -0.13719      0.03976      -3.45   0.00057 ***
## Year2009          -0.16834      0.03928      -4.29   1.9e-05 ***
## Year2010          -0.08500      0.04039      -2.10   0.03542 *
## Year2011          -0.07955      0.04072      -1.95   0.05085 .
## Year2012          -0.07683      0.03999      -1.92   0.05482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.314
## Multiple R-squared:  0.0263, Adjusted R-squared:  0.0198
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2289 is an outlier with |weight| = 0 ( < 3.9e-05);
## 223 weights are ~= 1. The remaining 2312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0848 0.8680 0.9520 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.94e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2536"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1310"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 477 456 433 410 390 522 464 384 403 427 412 424 367 363 344
## 2011 2012
## 325 323
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 145 141 123 138 99 146 166 179 180 223 225 214 193 182 165

```

```
## 2011 2012
## 164 172
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 137 124 113 110 76 128 136 133 141 165 190 187 165 156 139
## 2011 2012
## 135 151
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 73, df = 16, p-value = 3e-09
```

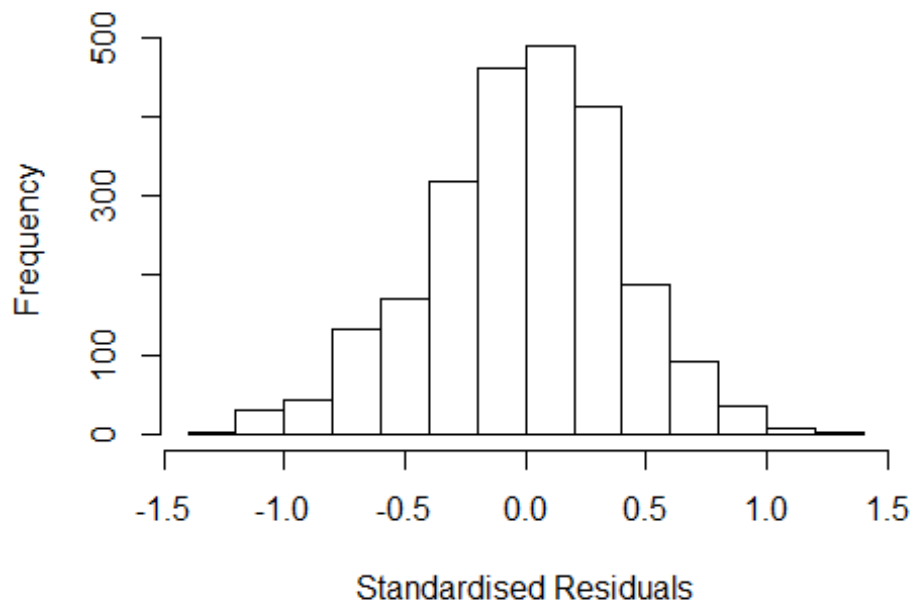


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.14, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 5.1808671679976"
## [1] "Male first author team size 2018 geometric mean: 4.48073123709671"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.14473370761185"
## [1] "Male last author team size 2018 geometric mean: 4.64512211317715"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.094 1      1.046
## LastAuthorFemale  1.051 1      1.025
## UniqueAuthors    1.244 4      1.028
## Year              1.380 16     1.010
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3672 -0.2634  0.0145  0.2551  1.2715
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.773876   0.060851  12.72  < 2e-16 ***
## FirstAuthorFemale1 0.000826   0.017257   0.05   0.962
## LastAuthorFemale1 0.009720   0.020716   0.47   0.639
## UniqueAuthors2    0.293744   0.043937   6.69  2.9e-11 ***
## UniqueAuthors3    0.296053   0.040725   7.27  4.9e-13 ***
## UniqueAuthors4    0.385403   0.040605   9.49  < 2e-16 ***
## UniqueAuthors5    0.465436   0.037104  12.54  < 2e-16 ***
## Year1997          0.118199   0.063886   1.85   0.064 .
## Year1998          0.058384   0.061052   0.96   0.339
## Year1999          0.059720   0.061792   0.97   0.334
```

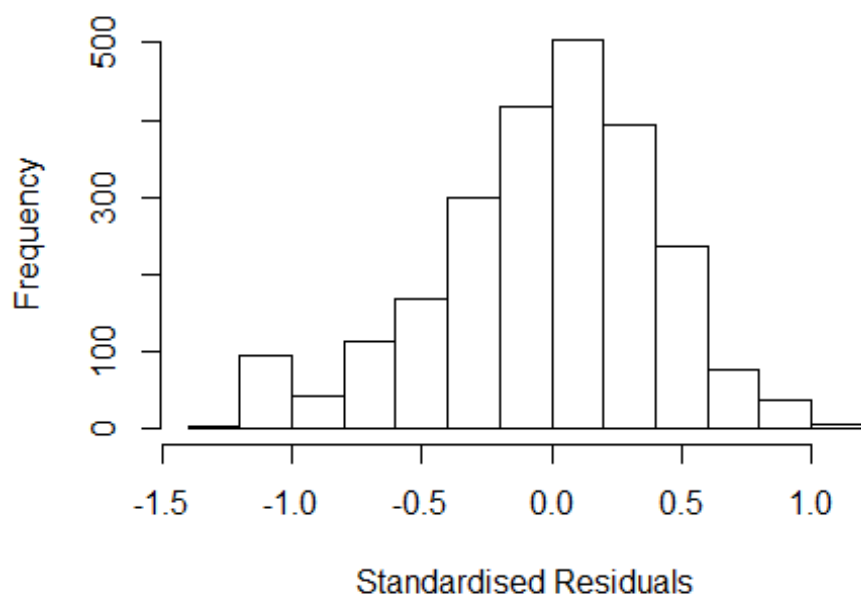
```

## Year2000      0.054405  0.070318  0.77  0.439
## Year2001     -0.037996  0.062015 -0.61  0.540
## Year2002     -0.088878  0.066243 -1.34  0.180
## Year2003     -0.107080  0.063727 -1.68  0.093
## Year2004     -0.094140  0.061284 -1.54  0.125
## Year2005     -0.084923  0.060493 -1.40  0.160
## Year2006     -0.087012  0.058826 -1.48  0.139
## Year2007     -0.076471  0.061043 -1.25  0.210
## Year2008     -0.009447  0.059722 -0.16  0.874
## Year2009     -0.012806  0.059447 -0.22  0.829
## Year2010     -0.021348  0.062898 -0.34  0.734
## Year2011     -0.042415  0.063410 -0.67  0.504
## Year2012     -0.007139  0.063938 -0.11  0.911
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.141, Adjusted R-squared:  0.133
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 203 weights are ~= 1. The remaining 2183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.164  0.867  0.950  0.896  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.075 1 1.037
## LastAuthorFemale 1.056 1 1.028
## Year 1.128 16 1.004

```



## Residuals from first and last author



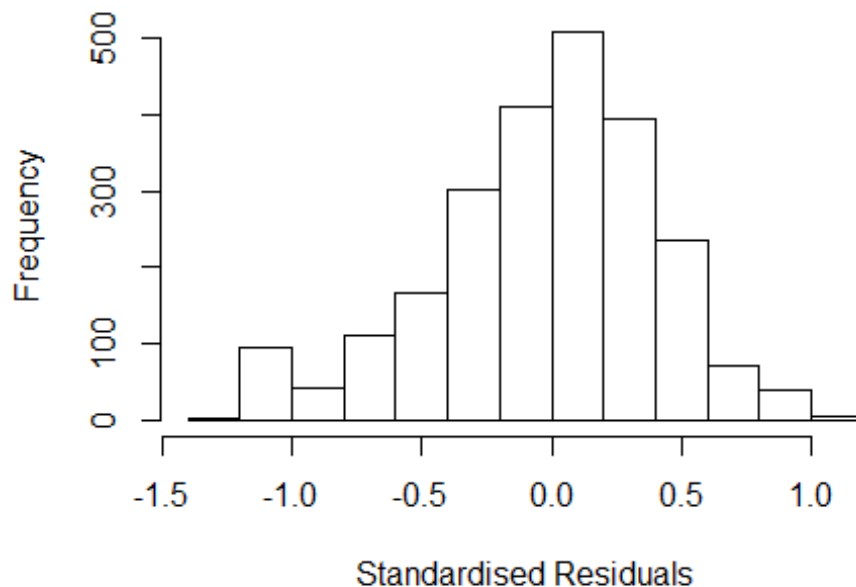
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2170 -0.2673  0.0224  0.2632  1.1803
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0873    0.0548   19.82  <2e-16 ***
## FirstAuthorFemale1  0.0215    0.0180    1.19   0.234
## LastAuthorFemale1 -0.0166    0.0230   -0.72   0.469
## Year1997          0.1464    0.0656    2.23   0.026 *
## Year1998          0.1014    0.0627    1.62   0.106
## Year1999          0.0456    0.0658    0.69   0.489
## Year2000          0.0709    0.0737    0.96   0.336
## Year2001         -0.0441    0.0665   -0.66   0.507
## Year2002         -0.0776    0.0707   -1.10   0.272
## Year2003         -0.0694    0.0692   -1.00   0.316
## Year2004         -0.0725    0.0651   -1.11   0.265
## Year2005         -0.0109    0.0614   -0.18   0.860
```

```

## Year2006          -0.0424      0.0611   -0.69    0.488
## Year2007          -0.0274      0.0628   -0.44    0.662
## Year2008           0.0465      0.0620    0.75    0.454
## Year2009           0.0496      0.0609    0.82    0.415
## Year2010           0.0566      0.0638    0.89    0.374
## Year2011           0.0167      0.0648    0.26    0.796
## Year2012           0.0644      0.0657    0.98    0.326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0226, Adjusted R-squared:  0.0151
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 189 weights are ~= 1. The remaining 2197 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.306  0.865   0.949   0.891   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"    "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.071 1      1.035
## Year              1.071 16      1.002

```

## Residuals from first author



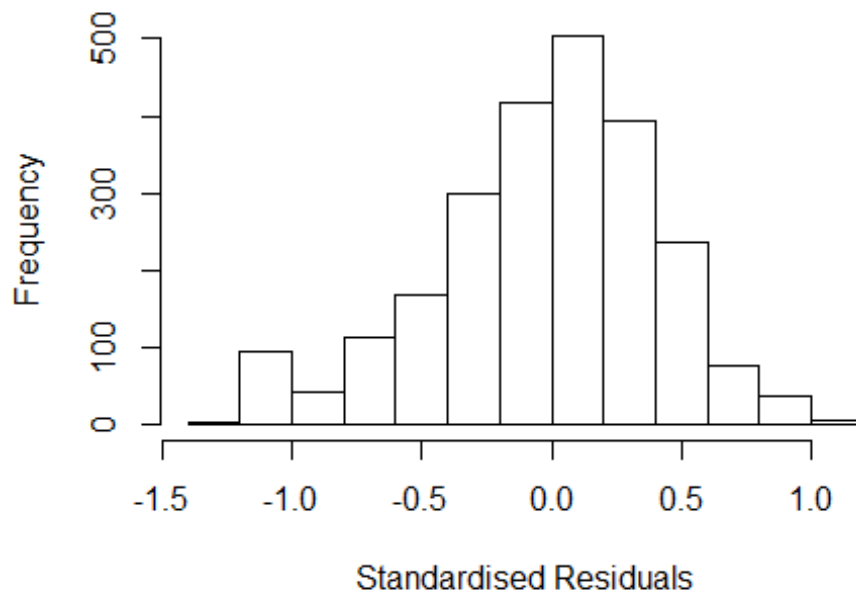
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2312 -0.2680 0.0233 0.2609 1.1835
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0845 0.0545 19.90 <2e-16 ***
## FirstAuthorFemale1 0.0196 0.0181 1.08 0.280
## Year1997 0.1467 0.0655 2.24 0.025 *
## Year1998 0.1015 0.0626 1.62 0.105
## Year1999 0.0460 0.0657 0.70 0.484
## Year2000 0.0725 0.0735 0.99 0.324
## Year2001 -0.0452 0.0665 -0.68 0.496
## Year2002 -0.0783 0.0708 -1.11 0.269
## Year2003 -0.0685 0.0691 -0.99 0.321
## Year2004 -0.0722 0.0650 -1.11 0.267
## Year2005 -0.0111 0.0613 -0.18 0.856
## Year2006 -0.0417 0.0610 -0.68 0.495
```

```

## Year2007          -0.0269      0.0627   -0.43    0.668
## Year2008          0.0464      0.0620    0.75    0.454
## Year2009          0.0494      0.0608    0.81    0.417
## Year2010          0.0570      0.0637    0.90    0.370
## Year2011          0.0156      0.0647    0.24    0.810
## Year2012          0.0640      0.0656    0.98    0.330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.0153
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 189 weights are ~= 1. The remaining 2197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.866  0.949  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.053 1          1.026
## Year            1.053 16          1.002

```

## Residuals from last author



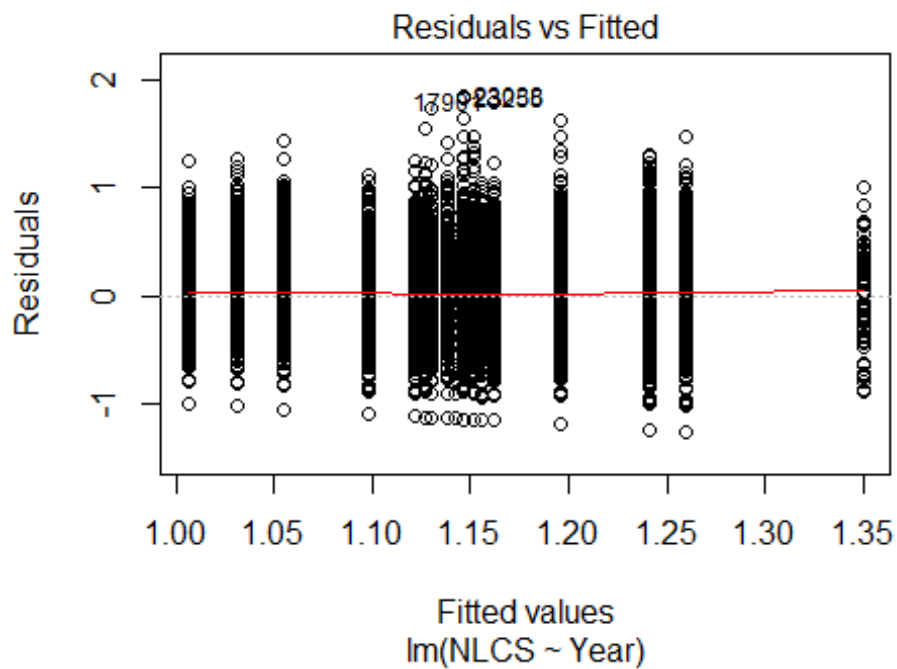
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2279 -0.2679  0.0223  0.2594  1.1710
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09333    0.05491   19.91  <2e-16 ***
## LastAuthorFemale1 -0.01303    0.02293   -0.57    0.570
## Year1997         0.14756    0.06582    2.24    0.025 *
## Year1998         0.09997    0.06291    1.59    0.112
## Year1999         0.04519    0.06598    0.68    0.493
## Year2000         0.07257    0.07384    0.98    0.326
## Year2001        -0.04314    0.06685   -0.65    0.519
## Year2002        -0.07625    0.07111   -1.07    0.284
## Year2003        -0.06603    0.06951   -0.95    0.342
## Year2004        -0.07119    0.06538   -1.09    0.276
## Year2005        -0.00877    0.06173   -0.14    0.887
## Year2006        -0.03968    0.06143   -0.65    0.518
```

```

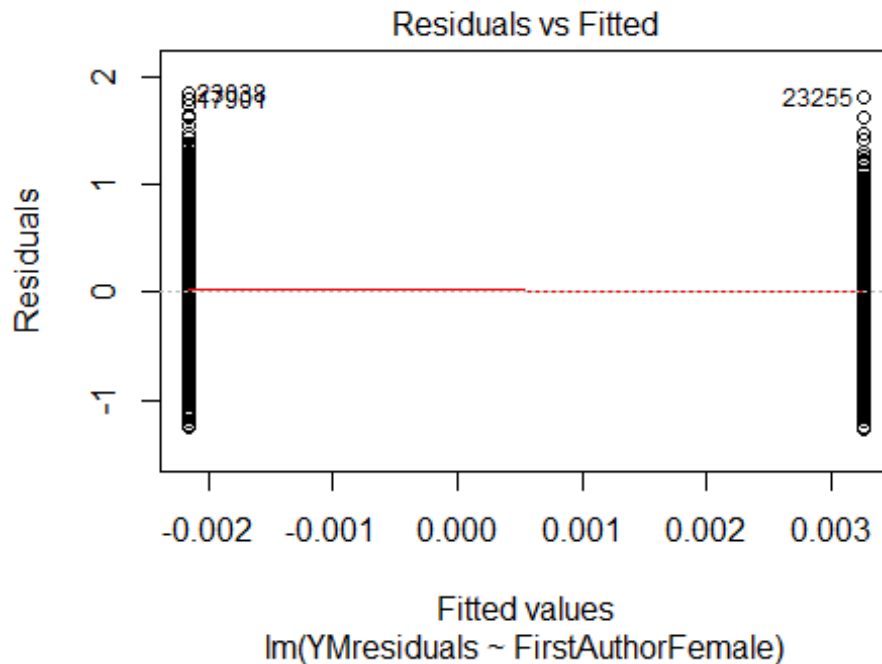
## Year2007          -0.02461    0.06310   -0.39    0.697
## Year2008          0.04935    0.06229    0.79    0.428
## Year2009          0.05279    0.06112    0.86    0.388
## Year2010          0.06130    0.06392    0.96    0.338
## Year2011          0.02144    0.06480    0.33    0.741
## Year2012          0.06764    0.06584    1.03    0.304
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.389
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.0149
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2201 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.297  0.863  0.949  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2386"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1311"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1525 1520 165 1209 150 1271 1153 1060 1174 1038 1012 1018 1035 1053 934
## 2011 2012
## 983 908
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 690 689 73 628 68 447 691 660 752 684 665 659 675 747 628
## 2011 2012

```

```
## 713 650
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 615 628 66 563 54 398 597 596 665 608 581 586 594 668 546
## 2011 2012
## 629 579
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```



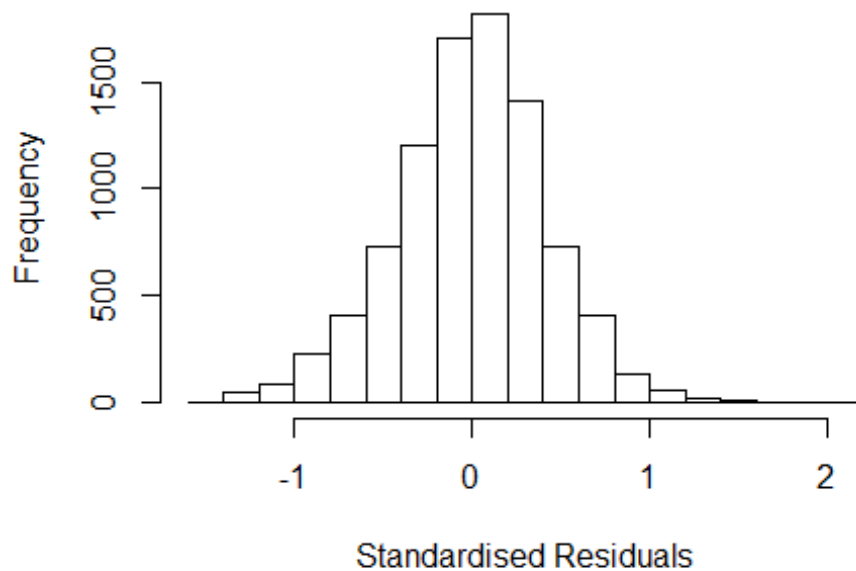
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 97, df = 1, p-value <2e-16
```



```
## [1] "Female first author team size 2018 geometric mean: 5.1918837944681"
## [1] "Male first author team size 2018 geometric mean: 4.09161846194151"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.2296468507575"
## [1] "Male last author team size 2018 geometric mean: 4.26971249230077"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27000, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1          1.015
## LastAuthorFemale  1.018 1          1.009
## UniqueAuthors    1.090 4          1.011
## Year              1.082 16         1.002
```



## Residuals from first and last author and team size



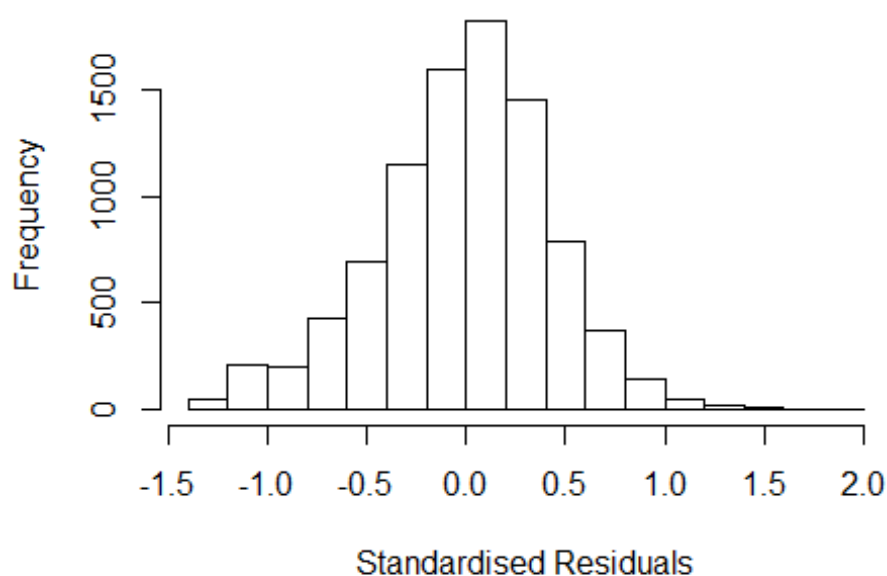
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40244 -0.26259 0.00946 0.26317 2.15473
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97559 0.02874 33.95 < 2e-16 ***
## FirstAuthorFemale1 -0.02608 0.00894 -2.92 0.0036 **
## LastAuthorFemale1 -0.01524 0.01069 -1.43 0.1541
## UniqueAuthors2 0.28575 0.02269 12.60 < 2e-16 ***
## UniqueAuthors3 0.32061 0.02220 14.44 < 2e-16 ***
## UniqueAuthors4 0.34768 0.02245 15.49 < 2e-16 ***
## UniqueAuthors5 0.41798 0.02085 20.04 < 2e-16 ***
## Year1997 0.00887 0.02780 0.32 0.7497
## Year1998 0.13212 0.06449 2.05 0.0405 *
## Year1999 -0.14620 0.02549 -5.74 1.0e-08 ***
```

```

## Year2000      -0.03288    0.05993   -0.55    0.5833
## Year2001      -0.16172    0.03008   -5.38    7.8e-08 ***
## Year2002      -0.17410    0.02781   -6.26    4.0e-10 ***
## Year2003      -0.24437    0.02626   -9.30    < 2e-16 ***
## Year2004      -0.23265    0.02618   -8.89    < 2e-16 ***
## Year2005      -0.17781    0.02579   -6.90    5.7e-12 ***
## Year2006      -0.15595    0.02650   -5.88    4.1e-09 ***
## Year2007      -0.15020    0.02545   -5.90    3.7e-09 ***
## Year2008      -0.12725    0.02617   -4.86    1.2e-06 ***
## Year2009      -0.14932    0.02613   -5.71    1.1e-08 ***
## Year2010      -0.16270    0.02703   -6.02    1.8e-09 ***
## Year2011      -0.13733    0.02617   -5.25    1.6e-07 ***
## Year2012      -0.10947    0.02722   -4.02    5.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.103, Adjusted R-squared:  0.101
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 8235 is an outlier with |weight| = 0 ( < 1.1e-05);
## 762 weights are ~= 1. The remaining 8210 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8620 0.9500 0.8970 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.11e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.008 1           1.004
## LastAuthorFemale 1.015 1           1.007
## Year 1.014 16           1.000

```

## Residuals from first and last author



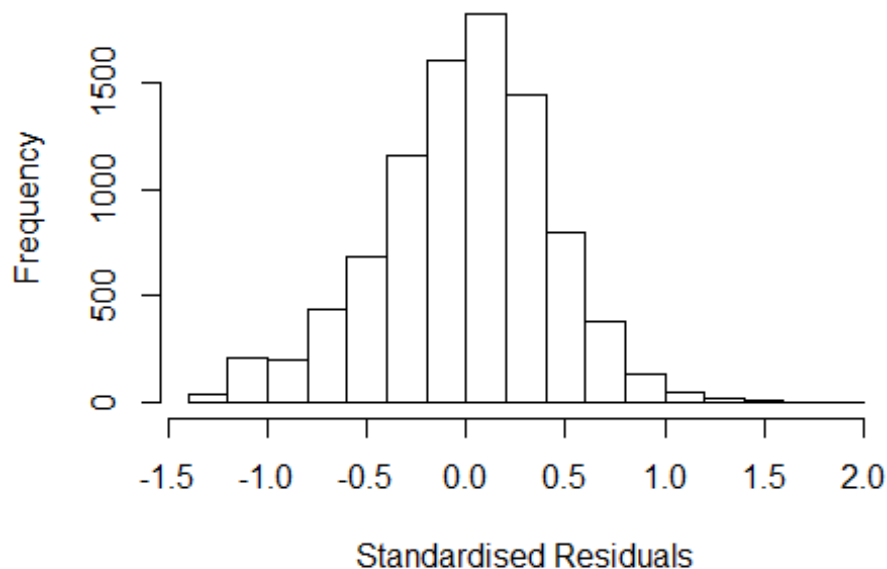
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2906 -0.2733  0.0165  0.2721  1.8345
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.27892    0.02142   59.70  < 2e-16 ***
## FirstAuthorFemale1  0.00627    0.00912    0.69   0.4918
## LastAuthorFemale1 -0.02218    0.01085   -2.04   0.0410 *
## Year1997         0.00538    0.02784    0.19   0.8468
## Year1998         0.07316    0.05739    1.27   0.2024
## Year1999        -0.12518    0.02582   -4.85  1.3e-06 ***
## Year2000        -0.06713    0.05810   -1.16   0.2479
## Year2001        -0.14076    0.03053   -4.61  4.1e-06 ***
## Year2002        -0.17824    0.02859   -6.23  4.7e-10 ***
## Year2003        -0.23853    0.02776   -8.59  < 2e-16 ***
## Year2004        -0.23340    0.02745   -8.50  < 2e-16 ***
## Year2005        -0.16078    0.02646   -6.08  1.3e-09 ***
```

```

## Year2006      -0.13157      0.02702      -4.87      1.1e-06 ***
## Year2007      -0.13070      0.02584      -5.06      4.3e-07 ***
## Year2008      -0.10939      0.02666      -4.10      4.1e-05 ***
## Year2009      -0.13207      0.02695      -4.90      9.7e-07 ***
## Year2010      -0.13780      0.02743      -5.02      5.2e-07 ***
## Year2011      -0.12042      0.02688      -4.48      7.5e-06 ***
## Year2012      -0.07963      0.02778      -2.87      0.0042 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0263, Adjusted R-squared:  0.0243
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 732 weights are ~= 1. The remaining 8241 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0031 0.8640 0.9500 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.004 1      1.002
## Year      1.004 16      1.000

```

## Residuals from first author



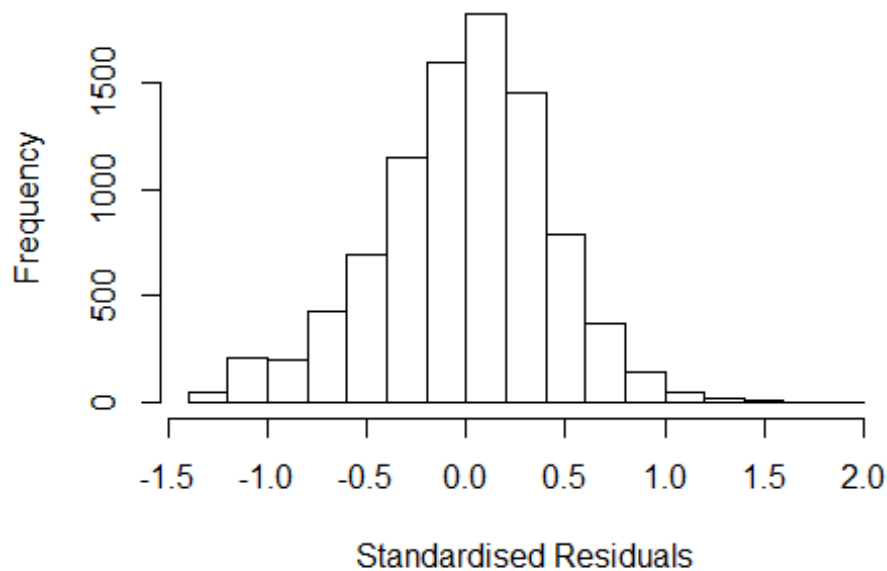
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2847 -0.2741 0.0178 0.2727 1.8384
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27494 0.02128 59.92 < 2e-16 ***
## FirstAuthorFemale1 0.00391 0.00912 0.43 0.6680
## Year1997 0.00584 0.02783 0.21 0.8338
## Year1998 0.07191 0.05760 1.25 0.2119
## Year1999 -0.12445 0.02582 -4.82 1.5e-06 ***
## Year2000 -0.06833 0.05825 -1.17 0.2408
## Year2001 -0.14095 0.03053 -4.62 3.9e-06 ***
## Year2002 -0.17751 0.02858 -6.21 5.5e-10 ***
## Year2003 -0.23820 0.02778 -8.57 < 2e-16 ***
## Year2004 -0.23277 0.02747 -8.47 < 2e-16 ***
## Year2005 -0.16077 0.02646 -6.08 1.3e-09 ***
## Year2006 -0.13110 0.02704 -4.85 1.3e-06 ***
```

```

## Year2007          -0.13051    0.02585   -5.05  4.5e-07 ***
## Year2008          -0.10952    0.02664   -4.11  4.0e-05 ***
## Year2009          -0.13218    0.02697   -4.90  9.7e-07 ***
## Year2010          -0.13792    0.02745   -5.02  5.1e-07 ***
## Year2011          -0.12034    0.02689   -4.47  7.8e-06 ***
## Year2012          -0.07988    0.02779   -2.87  0.0041 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.024
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 723 weights are ~= 1. The remaining 8250 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0027 0.8640 0.9500 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.01 1          1.005
## Year            1.01 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2866 -0.2732 0.0171 0.2718 1.8320
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28118 0.02101 60.98 < 2e-16 ***
## LastAuthorFemale1 -0.02121 0.01084 -1.96 0.0504 .
## Year1997 0.00543 0.02784 0.19 0.8455
## Year1998 0.07244 0.05742 1.26 0.2072
## Year1999 -0.12494 0.02583 -4.84 1.3e-06 ***
## Year2000 -0.06765 0.05814 -1.16 0.2446
## Year2001 -0.14061 0.03052 -4.61 4.1e-06 ***
## Year2002 -0.17834 0.02860 -6.24 4.7e-10 ***
## Year2003 -0.23841 0.02777 -8.58 < 2e-16 ***
## Year2004 -0.23331 0.02746 -8.50 < 2e-16 ***
## Year2005 -0.16068 0.02646 -6.07 1.3e-09 ***
## Year2006 -0.13137 0.02703 -4.86 1.2e-06 ***
```

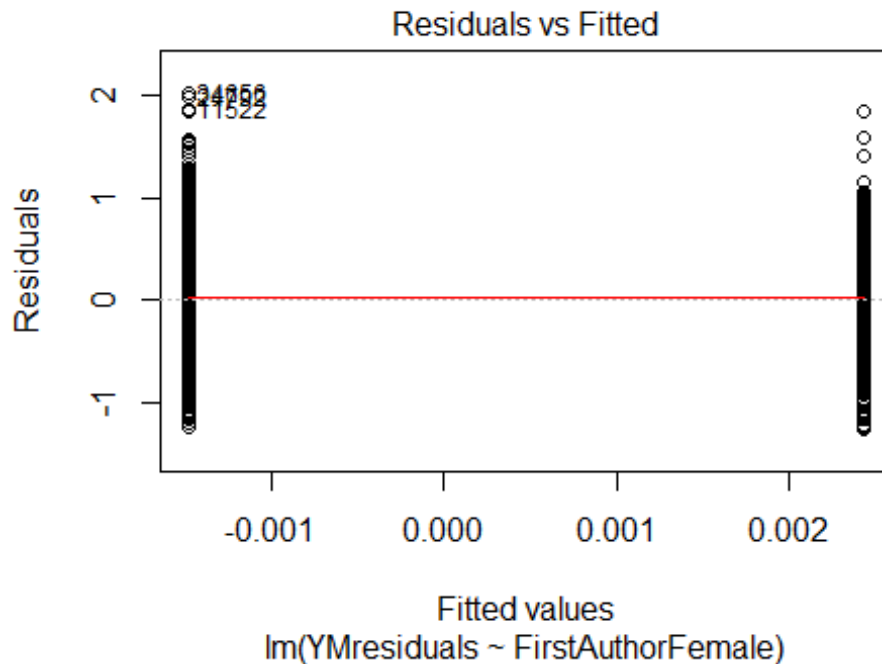
```

## Year2007          -0.13059      0.02584    -5.05  4.4e-07 ***
## Year2008          -0.10895      0.02666    -4.09  4.4e-05 ***
## Year2009          -0.13180      0.02696    -4.89  1.0e-06 ***
## Year2010          -0.13756      0.02744    -5.01  5.5e-07 ***
## Year2011          -0.12022      0.02689    -4.47  7.9e-06 ***
## Year2012          -0.07937      0.02778    -2.86  0.0043 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0263, Adjusted R-squared:  0.0244
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 736 weights are ~= 1. The remaining 8237 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0033 0.8640 0.9500 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 8973"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1312"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1884 1805 1806 1670 1858 1717 1746 1506 1524 1416 1401 1418 1549 1689 1585
## 2011 2012
## 1499 1292
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1134 1056 1031 1084 935 840 1283 1094 1176 1013 1006 1034 1121 1310 1176
## 2011 2012

```

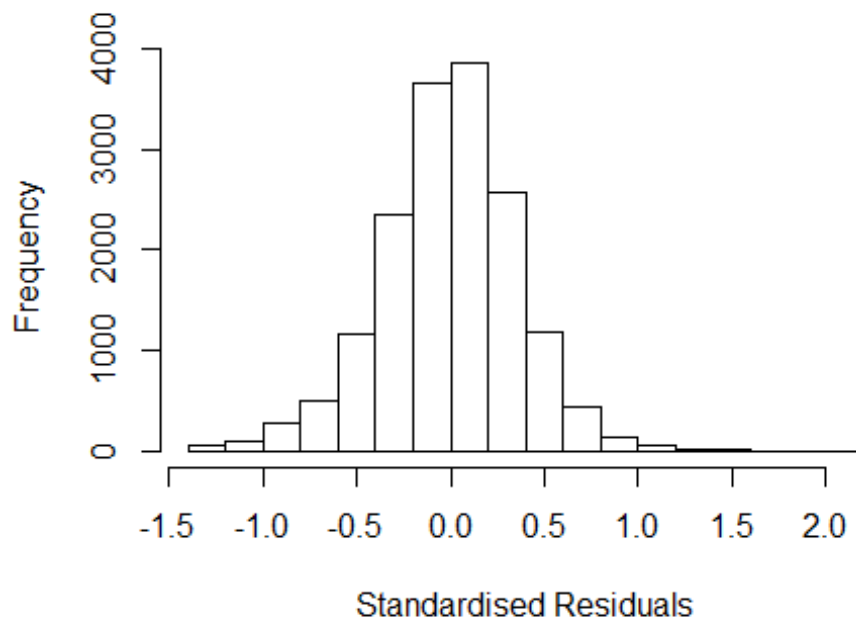






```
## [1] "Female first author team size 2018 geometric mean: 4.76149607688893"
## [1] "Male first author team size 2018 geometric mean: 4.26638466539222"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 61000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.75275439718624"
## [1] "Male last author team size 2018 geometric mean: 4.39503225051211"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 47000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022  1          1.011
## LastAuthorFemale  1.011  1          1.005
## UniqueAuthors    1.079  4          1.010
## Year              1.099 16          1.003
```

## Residuals from first and last author and team size



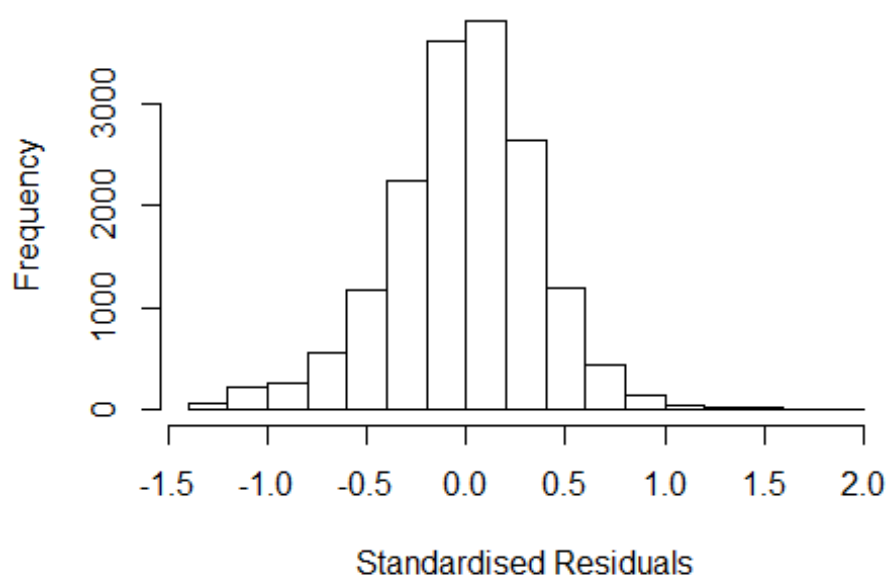
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37361 -0.22179  0.00619  0.22052  2.03685
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00661    0.02304   43.68 < 2e-16 ***
## FirstAuthorFemale1 -0.00570    0.00559   -1.02  0.3081
## LastAuthorFemale1 -0.00488    0.00704   -0.69  0.4882
## UniqueAuthors2     0.25017    0.01991   12.56 < 2e-16 ***
## UniqueAuthors3     0.26501    0.01959   13.53 < 2e-16 ***
## UniqueAuthors4     0.29486    0.01955   15.08 < 2e-16 ***
## UniqueAuthors5     0.35610    0.01893   18.82 < 2e-16 ***
## Year1997           0.01089    0.01936    0.56  0.5737
## Year1998          -0.05565    0.01824   -3.05  0.0023 **
## Year1999          -0.13914    0.01710   -8.14 4.3e-16 ***
```

```

## Year2000      -0.15469    0.01762   -8.78 < 2e-16 ***
## Year2001      -0.15626    0.01799   -8.69 < 2e-16 ***
## Year2002      -0.15521    0.01664   -9.33 < 2e-16 ***
## Year2003      -0.18960    0.01755  -10.80 < 2e-16 ***
## Year2004      -0.23381    0.01712  -13.66 < 2e-16 ***
## Year2005      -0.18876    0.01726  -10.94 < 2e-16 ***
## Year2006      -0.18292    0.01750  -10.45 < 2e-16 ***
## Year2007      -0.16243    0.01724   -9.42 < 2e-16 ***
## Year2008      -0.18144    0.01725  -10.52 < 2e-16 ***
## Year2009      -0.17143    0.01695  -10.11 < 2e-16 ***
## Year2010      -0.15256    0.01735   -8.79 < 2e-16 ***
## Year2011      -0.13763    0.01771   -7.77 8.2e-15 ***
## Year2012      -0.13578    0.01896   -7.16 8.4e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.331
## Multiple R-squared:  0.0834, Adjusted R-squared:  0.0822
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 10 observations
## c(3104,4690,4700,8400,10621,12961,13512,13991,14039,15092)
## are outliers with |weight| = 0 ( < 6.1e-06);
## 1344 weights are ~= 1. The remaining 15067 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0042 0.8640 0.9510 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      6.09e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
## factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1 1.009
## LastAuthorFemale 1.010 1 1.005
## Year 1.027 16 1.001

```

## Residuals from first and last author



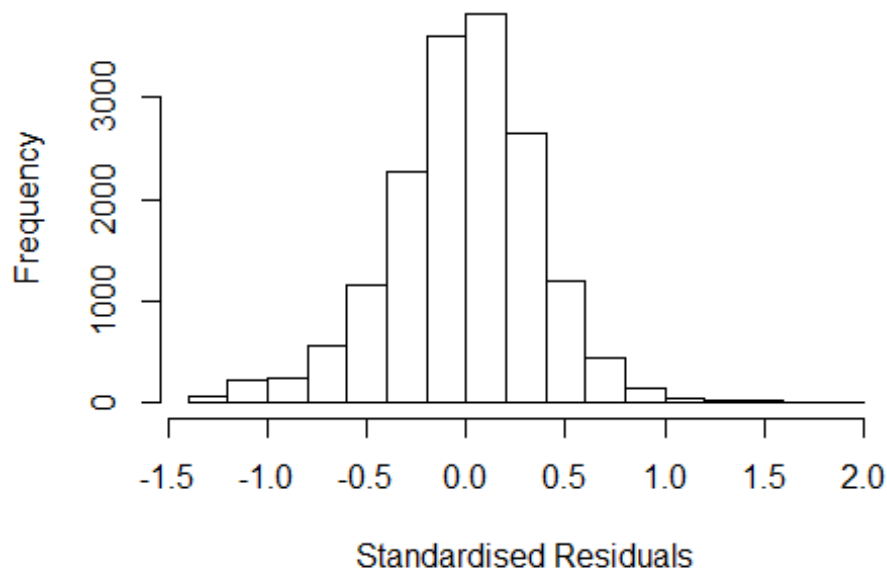
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29775 -0.22814 0.00457 0.22443 1.99206
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28392 0.01374 93.43 < 2e-16 ***
## FirstAuthorFemale1 0.00725 0.00568 1.28 0.2017
## LastAuthorFemale1 -0.00897 0.00717 -1.25 0.2107
## Year1997 0.00658 0.01955 0.34 0.7364
## Year1998 -0.05937 0.01832 -3.24 0.0012 **
## Year1999 -0.13627 0.01727 -7.89 3.2e-15 ***
## Year2000 -0.15178 0.01774 -8.56 < 2e-16 ***
## Year2001 -0.14909 0.01809 -8.24 < 2e-16 ***
## Year2002 -0.14369 0.01704 -8.43 < 2e-16 ***
## Year2003 -0.17407 0.01793 -9.71 < 2e-16 ***
## Year2004 -0.22574 0.01763 -12.81 < 2e-16 ***
## Year2005 -0.16906 0.01751 -9.65 < 2e-16 ***
```

```

## Year2006          -0.16238    0.01751   -9.27 < 2e-16 ***
## Year2007          -0.14222    0.01728   -8.23 < 2e-16 ***
## Year2008          -0.15853    0.01743   -9.10 < 2e-16 ***
## Year2009          -0.15591    0.01710   -9.12 < 2e-16 ***
## Year2010          -0.12950    0.01745   -7.42 1.2e-13 ***
## Year2011          -0.11998    0.01781   -6.74 1.7e-11 ***
## Year2012          -0.11232    0.01915   -5.87 4.5e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0268, Adjusted R-squared:  0.0258
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 5 observations c(4690,4700,10621,13991,15092)
## are outliers with |weight| = 0 ( < 6.1e-06);
## 1411 weights are ~= 1. The remaining 15005 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 0.8630 0.9500 0.8900 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.09e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1      1.009
## Year              1.018 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29624 -0.22779  0.00483  0.22355  1.99383
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.28268    0.01369   93.68 < 2e-16 ***
## FirstAuthorFemale1 0.00681    0.00569    1.20  0.2312
## Year1997          0.00675    0.01955    0.35  0.7298
## Year1998         -0.05943    0.01833   -3.24  0.0012 **
## Year1999         -0.13623    0.01728   -7.89 3.3e-15 ***
## Year2000         -0.15203    0.01775   -8.57 < 2e-16 ***
## Year2001         -0.14933    0.01809   -8.25 < 2e-16 ***
## Year2002         -0.14388    0.01705   -8.44 < 2e-16 ***
## Year2003         -0.17422    0.01793   -9.72 < 2e-16 ***
## Year2004         -0.22596    0.01763  -12.82 < 2e-16 ***
## Year2005         -0.16921    0.01751   -9.66 < 2e-16 ***
## Year2006         -0.16272    0.01752   -9.29 < 2e-16 ***
```

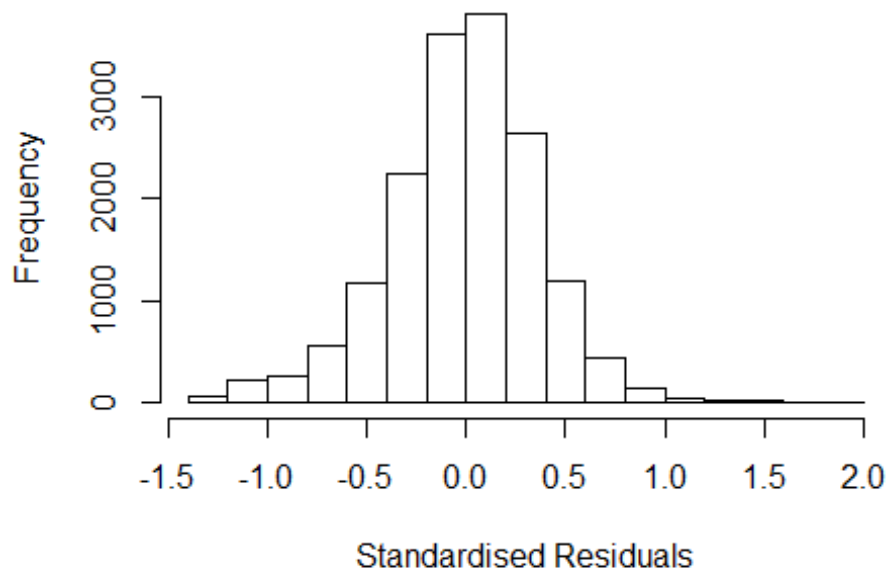
```

## Year2007          -0.14262    0.01728   -8.25 < 2e-16 ***
## Year2008          -0.15899    0.01742   -9.12 < 2e-16 ***
## Year2009          -0.15642    0.01710   -9.15 < 2e-16 ***
## Year2010          -0.12998    0.01745   -7.45 9.9e-14 ***
## Year2011          -0.12051    0.01782   -6.76 1.4e-11 ***
## Year2012          -0.11283    0.01914   -5.89 3.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.0257
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 5 observations c(4690,4700,10621,13991,15092)
## are outliers with |weight| = 0 ( < 6.1e-06);
## 1426 weights are ~= 1. The remaining 14990 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8640 0.9500 0.8900 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.09e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.01 1          1.005
## Year            1.01 16          1.000

```



## Residuals from last author



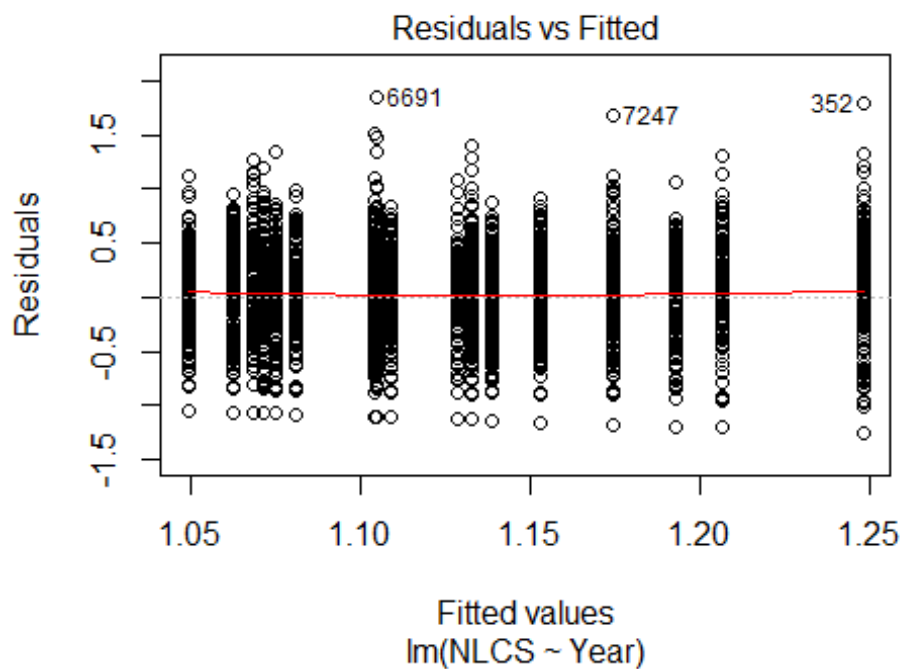
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29296 -0.22866  0.00559  0.22395  1.98897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28626    0.01359   94.63  < 2e-16 ***
## LastAuthorFemale1 -0.00841    0.00718   -1.17   0.2414
## Year1997         0.00670    0.01956    0.34   0.7320
## Year1998        -0.05957    0.01832   -3.25   0.0012 **
## Year1999        -0.13583    0.01726   -7.87  3.8e-15 ***
## Year2000        -0.15167    0.01774   -8.55  < 2e-16 ***
## Year2001        -0.14874    0.01808   -8.23  < 2e-16 ***
## Year2002        -0.14350    0.01705   -8.42  < 2e-16 ***
## Year2003        -0.17357    0.01792   -9.69  < 2e-16 ***
## Year2004        -0.22521    0.01762  -12.78  < 2e-16 ***
## Year2005        -0.16819    0.01749   -9.62  < 2e-16 ***
## Year2006        -0.16186    0.01751   -9.24  < 2e-16 ***
```

```

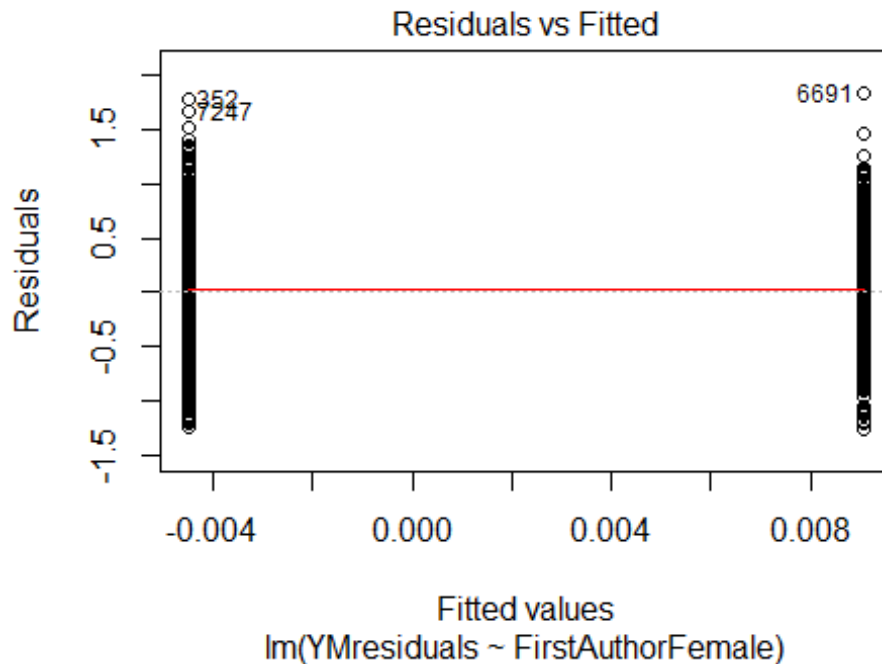
## Year2007          -0.14185      0.01728      -8.21   2.4e-16 ***
## Year2008          -0.15802      0.01741      -9.07   < 2e-16 ***
## Year2009          -0.15538      0.01709      -9.09   < 2e-16 ***
## Year2010          -0.12876      0.01743      -7.39   1.6e-13 ***
## Year2011          -0.11923      0.01779      -6.70   2.1e-11 ***
## Year2012          -0.11148      0.01912      -5.83   5.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.337
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.0257
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 5 observations c(4690,4700,10621,13991,15092)
## are outliers with |weight| = 0 ( < 6.1e-06);
## 1404 weights are ~= 1. The remaining 15012 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0005 0.8630 0.9500 0.8900 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.09e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 16421"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1313"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 293 234 229 240 294 314 273 256 304 294 390 342 353 375 393
## 2011 2012
## 434 420
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 185 137 127 160 144 166 177 181 214 210 294 253 281 288 286
## 2011 2012
## 333 298
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 169 121 118 148 132 150 161 163 196 181 259 222 247 257 257
## 2011 2012
## 307 258
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```

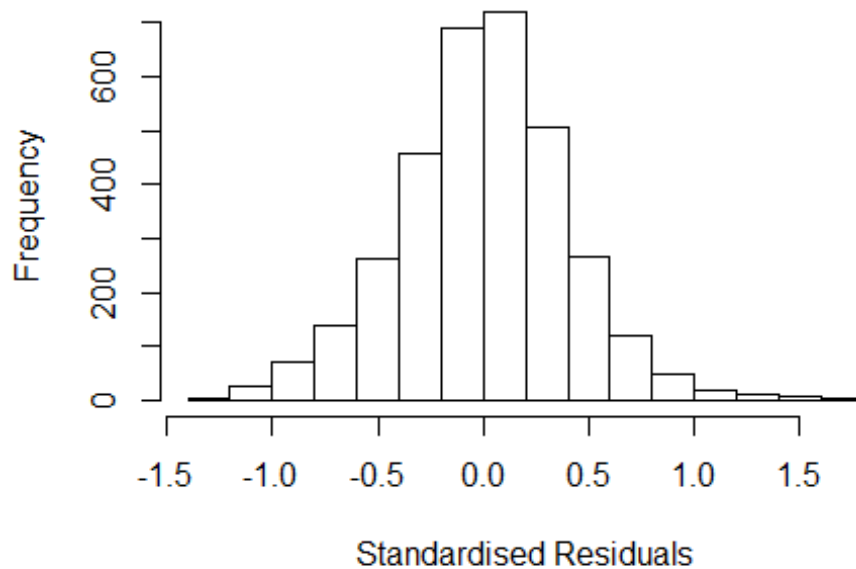


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 5.24046363285639"
## [1] "Male first author team size 2018 geometric mean: 5.69250712862708"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.28162136805888"
## [1] "Male last author team size 2018 geometric mean: 5.53762218372527"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.064 1          1.031
## LastAuthorFemale  1.027 1          1.014
## UniqueAuthors    1.151 4          1.018
## Year             1.193 16          1.006
```

## Residuals from first and last author and team size



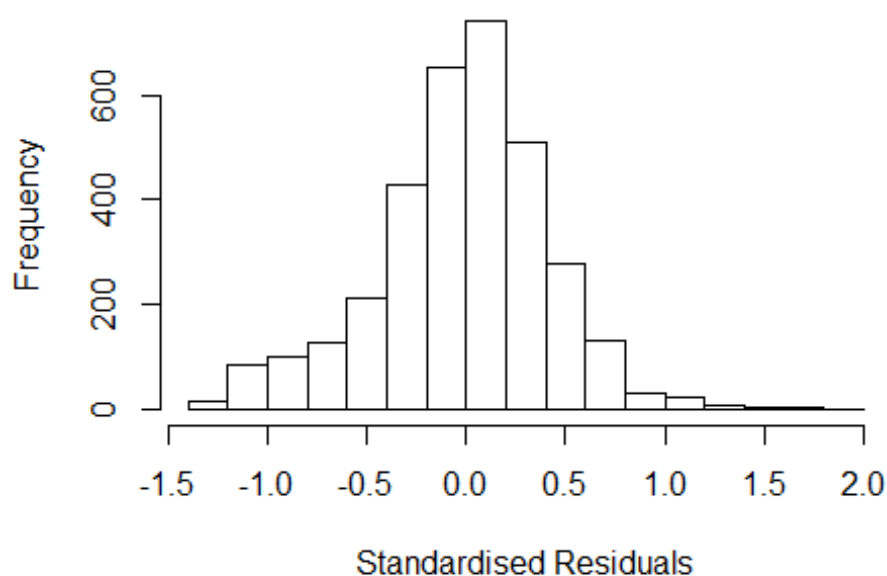
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30502 -0.24293  0.00719  0.25174  1.76844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.79484    0.05735   13.86 < 2e-16 ***
## FirstAuthorFemale1 -0.00972    0.01468   -0.66  0.50775
## LastAuthorFemale1 -0.03131    0.01713   -1.83  0.06762 .
## UniqueAuthors2     0.27435    0.05018    5.47 4.9e-08 ***
## UniqueAuthors3     0.46655    0.04390   10.63 < 2e-16 ***
## UniqueAuthors4     0.44750    0.04275   10.47 < 2e-16 ***
## UniqueAuthors5     0.55655    0.03876   14.36 < 2e-16 ***
## Year1997          -0.12182    0.07377   -1.65  0.09875 .
## Year1998          -0.00891    0.06940   -0.13  0.89788
## Year1999          -0.03665    0.05321   -0.69  0.49105
```

```

## Year2000      -0.12484      0.05470      -2.28      0.02255      *
## Year2001      -0.13919      0.05589      -2.49      0.01281      *
## Year2002      -0.14955      0.05291      -2.83      0.00474      **
## Year2003      -0.18340      0.05409      -3.39      0.00071      ***
## Year2004      -0.12851      0.05262      -2.44      0.01466      *
## Year2005      -0.15001      0.05116      -2.93      0.00339      **
## Year2006      -0.14442      0.05026      -2.87      0.00409      **
## Year2007      -0.14945      0.04919      -3.04      0.00240      **
## Year2008      -0.18132      0.04968      -3.65      0.00027      ***
## Year2009      -0.12841      0.04935      -2.60      0.00930      **
## Year2010      -0.12521      0.04904      -2.55      0.01072      *
## Year2011      -0.15911      0.04948      -3.22      0.00131      **
## Year2012      -0.09376      0.05006      -1.87      0.06116      .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.369
## Multiple R-squared:  0.158, Adjusted R-squared:  0.153
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 2994 is an outlier with |weight| = 0 ( < 3e-05);
## 278 weights are ~= 1. The remaining 3067 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0026 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max      maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1      1.019
## LastAuthorFemale 1.029 1      1.015
## Year      1.067 16      1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2628 -0.2573 0.0143 0.2510 1.8305
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2628 0.0487 25.94 < 2e-16 ***
## FirstAuthorFemale1 0.0116 0.0152 0.76 0.44696
## LastAuthorFemale1 -0.0233 0.0182 -1.28 0.20127
## Year1997 -0.1520 0.0827 -1.84 0.06597 .
## Year1998 -0.0212 0.0721 -0.29 0.76894
## Year1999 -0.0398 0.0582 -0.68 0.49411
## Year2000 -0.1148 0.0567 -2.02 0.04295 *
## Year2001 -0.1719 0.0588 -2.92 0.00349 **
## Year2002 -0.1618 0.0583 -2.77 0.00555 **
## Year2003 -0.2010 0.0591 -3.40 0.00068 ***
## Year2004 -0.1510 0.0577 -2.62 0.00894 **
## Year2005 -0.1520 0.0558 -2.73 0.00644 **
```

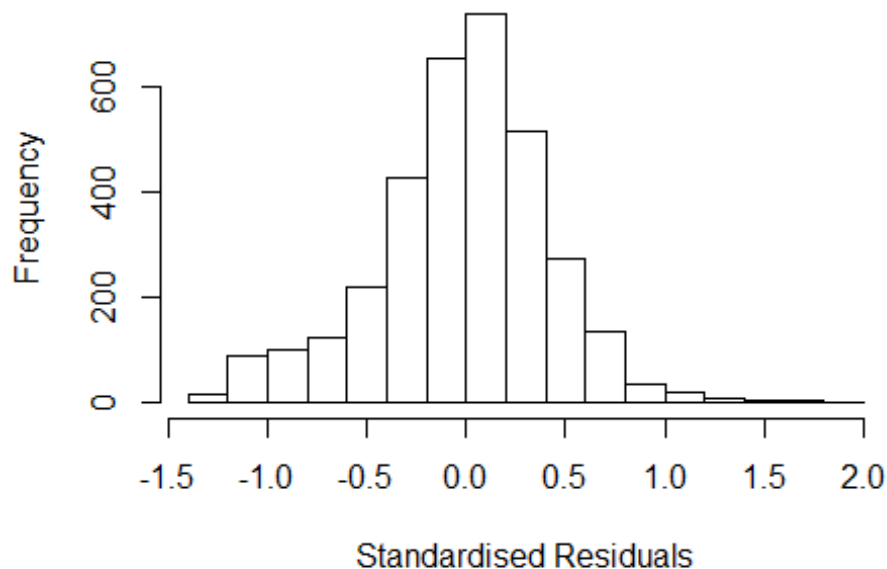
```

## Year2006          -0.1277      0.0551   -2.32  0.02057 *
## Year2007          -0.1367      0.0538   -2.54  0.01105 *
## Year2008          -0.1786      0.0544   -3.28  0.00104 **
## Year2009          -0.1118      0.0531   -2.11  0.03519 *
## Year2010          -0.1071      0.0532   -2.01  0.04399 *
## Year2011          -0.1539      0.0539   -2.85  0.00435 **
## Year2012          -0.0840      0.0543   -1.55  0.12153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.0103
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(169,2994) are outliers with |weight| = 0 ( < 3e-05);
## 278 weights are ~ = 1. The remaining 3066 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0131 0.8580 0.9500 0.8850 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1          1.019
## Year              1.039 16          1.001

```



## Residuals from first author



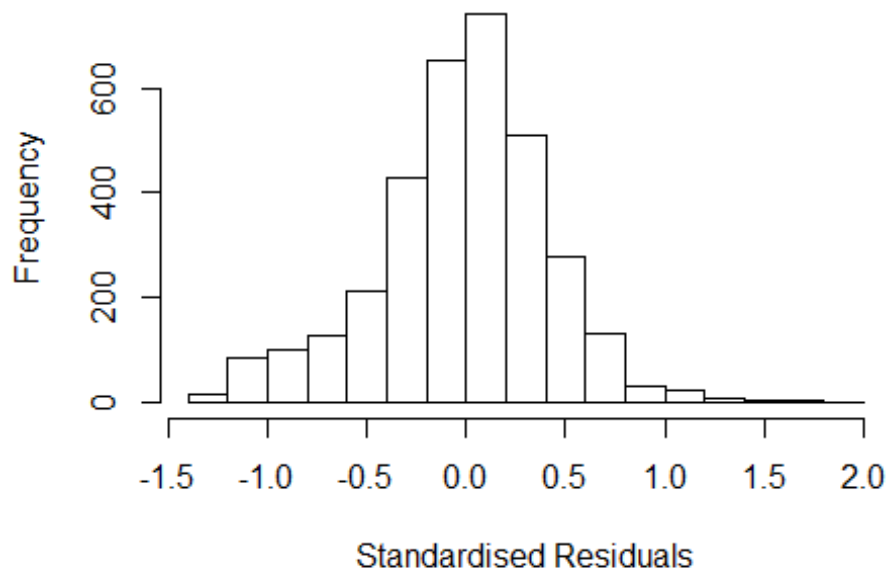
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.270 -0.257 0.014 0.251 1.837
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2596 0.0486 25.90 < 2e-16 ***
## FirstAuthorFemale1 0.0106 0.0153 0.70 0.48660
## Year1997 -0.1520 0.0823 -1.85 0.06476 .
## Year1998 -0.0203 0.0721 -0.28 0.77826
## Year1999 -0.0402 0.0582 -0.69 0.48958
## Year2000 -0.1144 0.0568 -2.01 0.04401 *
## Year2001 -0.1716 0.0589 -2.92 0.00358 **
## Year2002 -0.1623 0.0583 -2.78 0.00543 **
## Year2003 -0.2007 0.0592 -3.39 0.00071 ***
## Year2004 -0.1514 0.0578 -2.62 0.00881 **
## Year2005 -0.1525 0.0558 -2.73 0.00631 **
## Year2006 -0.1293 0.0551 -2.35 0.01898 *
```

```

## Year2007          -0.1380      0.0537   -2.57  0.01030 *
## Year2008          -0.1804      0.0544   -3.31  0.00093 ***
## Year2009          -0.1139      0.0530   -2.15  0.03175 *
## Year2010          -0.1086      0.0532   -2.04  0.04107 *
## Year2011          -0.1561      0.0539   -2.90  0.00379 **
## Year2012          -0.0859      0.0542   -1.59  0.11288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.0101
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(169,2994) are outliers with |weight| = 0 ( < 3e-05);
## 289 weights are ~= 1. The remaining 3055 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0118 0.8590 0.9500 0.8850 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1          1.014
## Year          1.028 16          1.001

```

## Residuals from last author



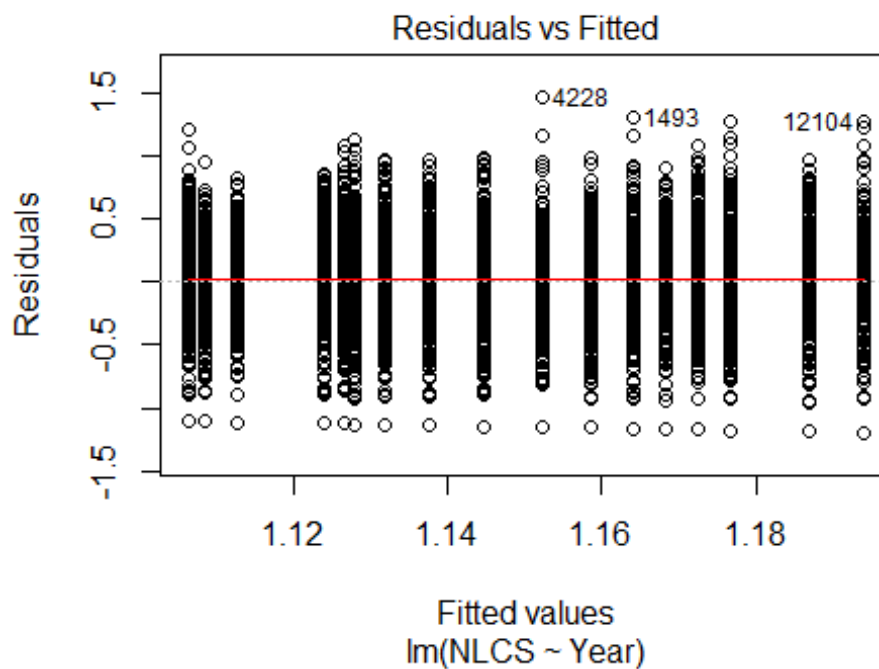
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2659 -0.2570 0.0143 0.2496 1.8382
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2659 0.0485 26.09 < 2e-16 ***
## LastAuthorFemale1 -0.0226 0.0183 -1.24 0.21575
## Year1997 -0.1520 0.0829 -1.83 0.06682 .
## Year1998 -0.0203 0.0721 -0.28 0.77795
## Year1999 -0.0403 0.0583 -0.69 0.48952
## Year2000 -0.1139 0.0568 -2.01 0.04493 *
## Year2001 -0.1716 0.0589 -2.91 0.00363 **
## Year2002 -0.1624 0.0585 -2.78 0.00551 **
## Year2003 -0.2009 0.0593 -3.39 0.00071 ***
## Year2004 -0.1512 0.0579 -2.61 0.00901 **
## Year2005 -0.1511 0.0559 -2.71 0.00686 **
## Year2006 -0.1275 0.0552 -2.31 0.02099 *
```

```

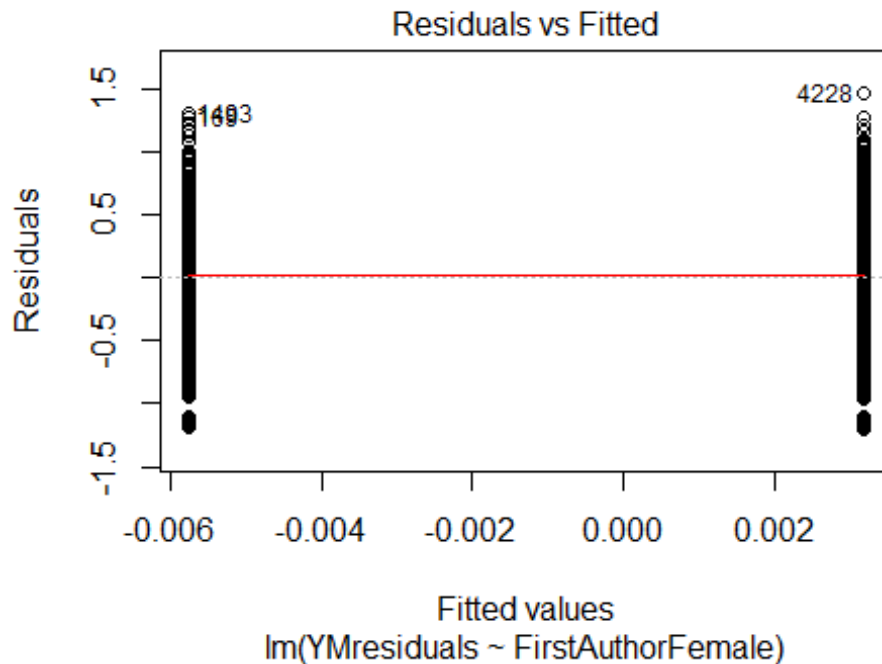
## Year2007          -0.1359      0.0539   -2.52  0.01175 *
## Year2008          -0.1778      0.0545   -3.26  0.00113 **
## Year2009          -0.1104      0.0532   -2.08  0.03791 *
## Year2010          -0.1056      0.0533   -1.98  0.04766 *
## Year2011          -0.1531      0.0541   -2.83  0.00467 **
## Year2012          -0.0819      0.0544   -1.51  0.13204
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0155, Adjusted R-squared:  0.0105
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(169,2994) are outliers with |weight| <= 4.9e-06 ( < 3e-
05);
## 279 weights are ~ = 1. The remaining 3065 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0142 0.8580 0.9510 0.8850 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3346"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1314"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 933 855 834 764 775 651 683 476 518 538 526 603 534 538 481
## 2011 2012
## 505 527
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 367 350 374 357 268 224 379 258 319 327 347 395 346 393 339
## 2011 2012
## 361 364
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 331 305 339 318 243 208 333 230 290 297 313 357 314 357 310
## 2011 2012
## 322 331
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.1
```

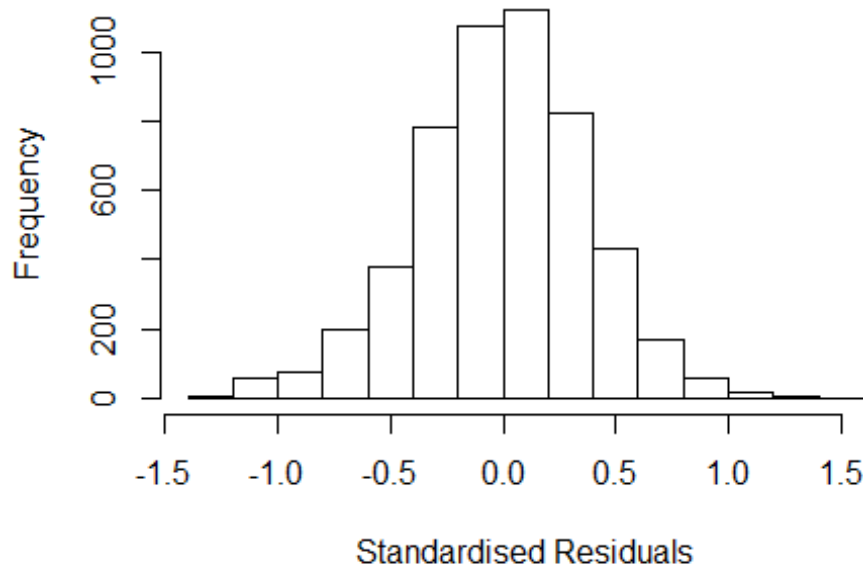


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.18846955335636"
## [1] "Male first author team size 2018 geometric mean: 3.70319177189224"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.26288931347895"
## [1] "Male last author team size 2018 geometric mean: 3.80231663953571"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1          1.014
## LastAuthorFemale  1.030 1          1.015
## UniqueAuthors     1.133 4          1.016
## Year              1.161 16         1.005
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32896 -0.24183  0.00456  0.24173  1.51828
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0618    0.0338   31.41 < 2e-16 ***
## FirstAuthorFemale1 -0.0216    0.0110   -1.96  0.0503 .
## LastAuthorFemale1 -0.0154    0.0134   -1.15  0.2503
## UniqueAuthors2     0.0749    0.0305    2.46  0.0140 *
## UniqueAuthors3     0.1618    0.0301    5.38 7.9e-08 ***
## UniqueAuthors4     0.1678    0.0310    5.41 6.5e-08 ***
## UniqueAuthors5     0.2541    0.0299    8.49 < 2e-16 ***
## Year1997          -0.0235    0.0321   -0.73  0.4633
## Year1998           0.0130    0.0316    0.41  0.6798
## Year1999          -0.0241    0.0307   -0.78  0.4329
```

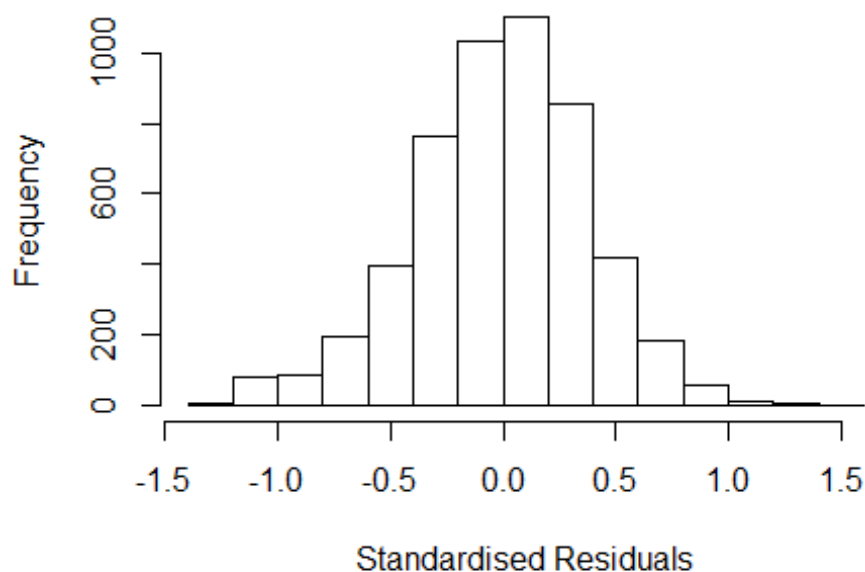
```

## Year2000          -0.0380      0.0331   -1.15   0.2512
## Year2001          -0.0892      0.0339   -2.63   0.0085 **
## Year2002          -0.0700      0.0308   -2.27   0.0230 *
## Year2003          -0.0383      0.0348   -1.10   0.2714
## Year2004          -0.1051      0.0337   -3.12   0.0018 **
## Year2005          -0.0817      0.0318   -2.57   0.0101 *
## Year2006          -0.0935      0.0312   -2.99   0.0028 **
## Year2007          -0.0899      0.0301   -2.99   0.0028 **
## Year2008          -0.0686      0.0316   -2.17   0.0301 *
## Year2009          -0.0293      0.0300   -0.98   0.3293
## Year2010          -0.0213      0.0304   -0.70   0.4841
## Year2011          -0.0702      0.0332   -2.11   0.0345 *
## Year2012          -0.0799      0.0326   -2.45   0.0143 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.0486, Adjusted R-squared:  0.0445
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 427 weights are ~= 1. The remaining 4771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.037  0.867  0.951  0.898  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.92e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1 1.009
## LastAuthorFemale 1.027 1 1.013
## Year 1.033 16 1.001

```



## Residuals from first and last author



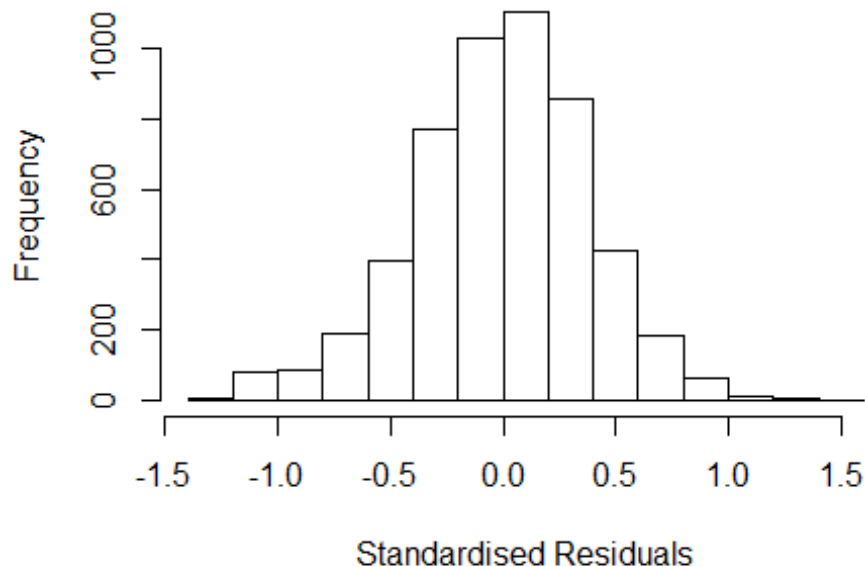
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21015 -0.24932  0.00698  0.24649  1.44760
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.18502    0.02365   50.11  <2e-16 ***
## FirstAuthorFemale1 -0.01258    0.01116   -1.13    0.260
## LastAuthorFemale1  -0.01844    0.01366   -1.35    0.177
## Year1997          -0.00814    0.03227   -0.25    0.801
## Year1998           0.02513    0.03195    0.79    0.432
## Year1999          -0.01030    0.03060   -0.34    0.737
## Year2000          -0.01563    0.03352   -0.47    0.641
## Year2001          -0.05607    0.03417   -1.64    0.101
## Year2002          -0.03563    0.03144   -1.13    0.257
## Year2003           0.00259    0.03486    0.07    0.941
## Year2004          -0.06284    0.03422   -1.84    0.066 .
## Year2005          -0.04204    0.03187   -1.32    0.187
```

```

## Year2006      -0.05380    0.03126   -1.72    0.085 .
## Year2007      -0.04709    0.03038   -1.55    0.121
## Year2008      -0.01886    0.03217   -0.59    0.558
## Year2009       0.01460    0.03016    0.48    0.628
## Year2010       0.02377    0.03046    0.78    0.435
## Year2011      -0.02122    0.03333   -0.64    0.524
## Year2012      -0.03095    0.03276   -0.94    0.345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.00575,    Adjusted R-squared:  0.00229
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 465 weights are ~= 1. The remaining 4733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0846 0.8690 0.9500 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1      1.006
## Year      1.012 16      1.000

```

## Residuals from first author



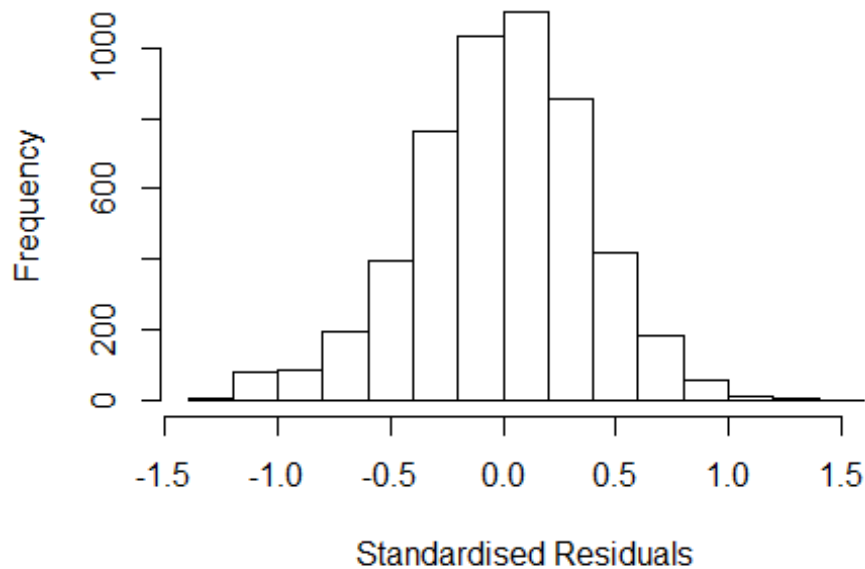
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20675 -0.24928  0.00647  0.24675  1.45014
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.18311    0.02358   50.16  <2e-16 ***
## FirstAuthorFemale1 -0.01399    0.01113   -1.26    0.209
## Year1997         -0.00875    0.03225   -0.27    0.786
## Year1998          0.02364    0.03193    0.74    0.459
## Year1999         -0.01036    0.03059   -0.34    0.735
## Year2000         -0.01625    0.03352   -0.48    0.628
## Year2001         -0.05612    0.03418   -1.64    0.101
## Year2002         -0.03586    0.03146   -1.14    0.254
## Year2003          0.00119    0.03486    0.03    0.973
## Year2004         -0.06321    0.03423   -1.85    0.065 .
## Year2005         -0.04259    0.03189   -1.34    0.182
## Year2006         -0.05564    0.03123   -1.78    0.075 .
```

```

## Year2007          -0.04891    0.03038   -1.61    0.108
## Year2008          -0.02057    0.03220   -0.64    0.523
## Year2009           0.01288    0.03016    0.43    0.669
## Year2010           0.02212    0.03046    0.73    0.468
## Year2011          -0.02339    0.03335   -0.70    0.483
## Year2012          -0.03317    0.03274   -1.01    0.311
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.00536,    Adjusted R-squared:  0.0021
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 462 weights are ~= 1. The remaining 4736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.083  0.870  0.950  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.022 1          1.011
## Year            1.022 16          1.001

```

## Residuals from last author



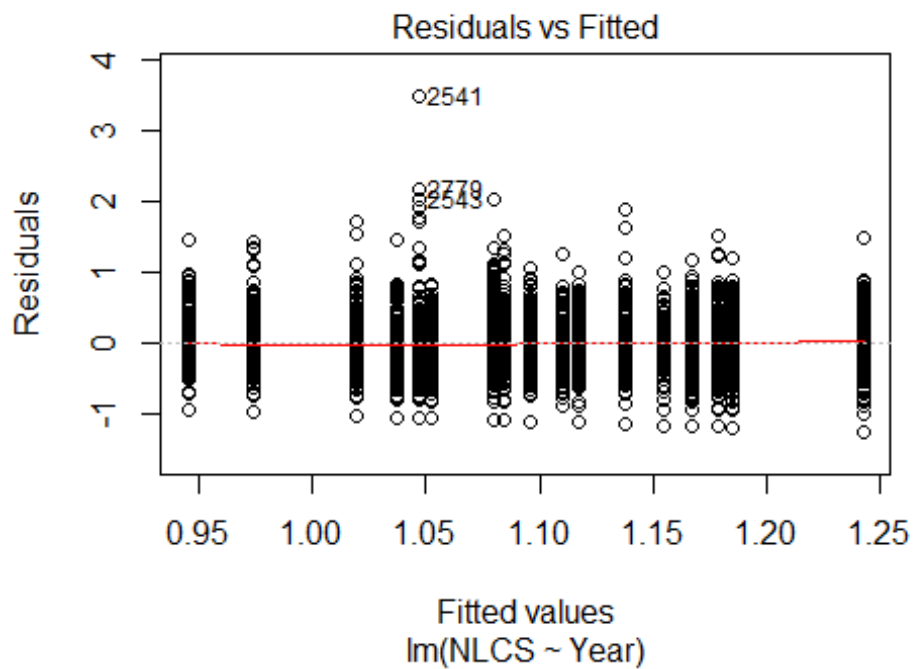
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20648 -0.24965 0.00622 0.24664 1.45139
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18128 0.02336 50.57 <2e-16 ***
## LastAuthorFemale1 -0.01987 0.01361 -1.46 0.144
## Year1997 -0.00844 0.03229 -0.26 0.794
## Year1998 0.02520 0.03196 0.79 0.430
## Year1999 -0.01093 0.03061 -0.36 0.721
## Year2000 -0.01567 0.03351 -0.47 0.640
## Year2001 -0.05657 0.03414 -1.66 0.098 .
## Year2002 -0.03515 0.03142 -1.12 0.263
## Year2003 0.00232 0.03484 0.07 0.947
## Year2004 -0.06360 0.03418 -1.86 0.063 .
## Year2005 -0.04222 0.03185 -1.33 0.185
## Year2006 -0.05404 0.03123 -1.73 0.084 .
```

```

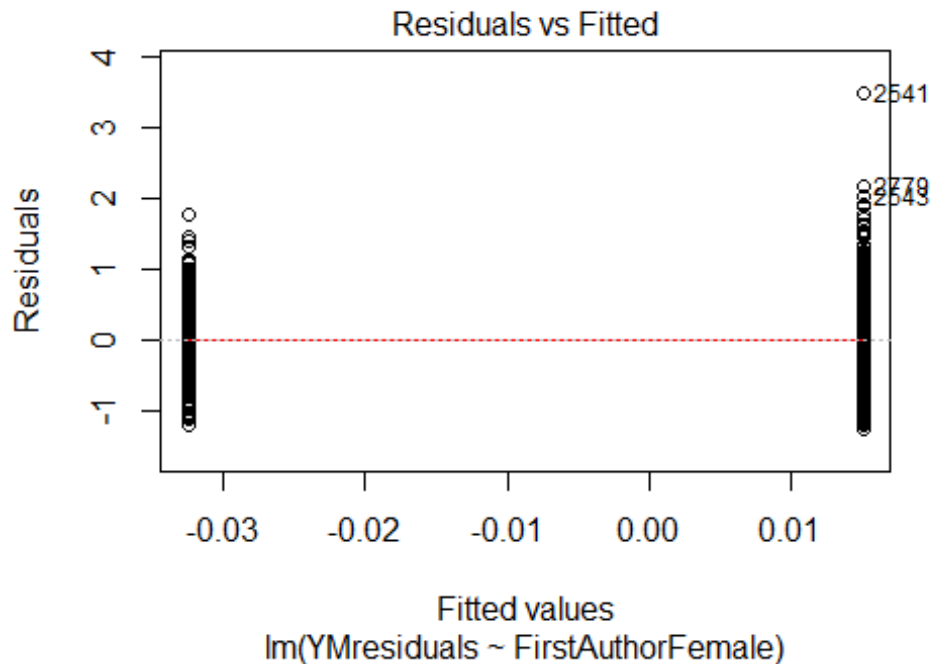
## Year2007          -0.04833      0.03029      -1.60      0.111
## Year2008          -0.01934      0.03212      -0.60      0.547
## Year2009           0.01377      0.03014       0.46      0.648
## Year2010           0.02336      0.03046       0.77      0.443
## Year2011          -0.02282      0.03328      -0.69      0.493
## Year2012          -0.03216      0.03276      -0.98      0.326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0055, Adjusted R-squared:  0.00224
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 469 weights are ~= 1. The remaining 4729 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0822 0.8700 0.9500 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5198"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1315"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 250 215 271 257 219 230 206 235 233 250 219 262 213 230 237
## 2011 2012
## 201 175
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 175 155 179 173 144 114 146 181 191 192 174 212 156 180 180
## 2011 2012

```

```
## 166 135
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 151 139 165 157 128 103 138 160 167 181 158 187 129 164 163
## 2011 2012
## 144 120
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 90, df = 16, p-value = 2e-12
```



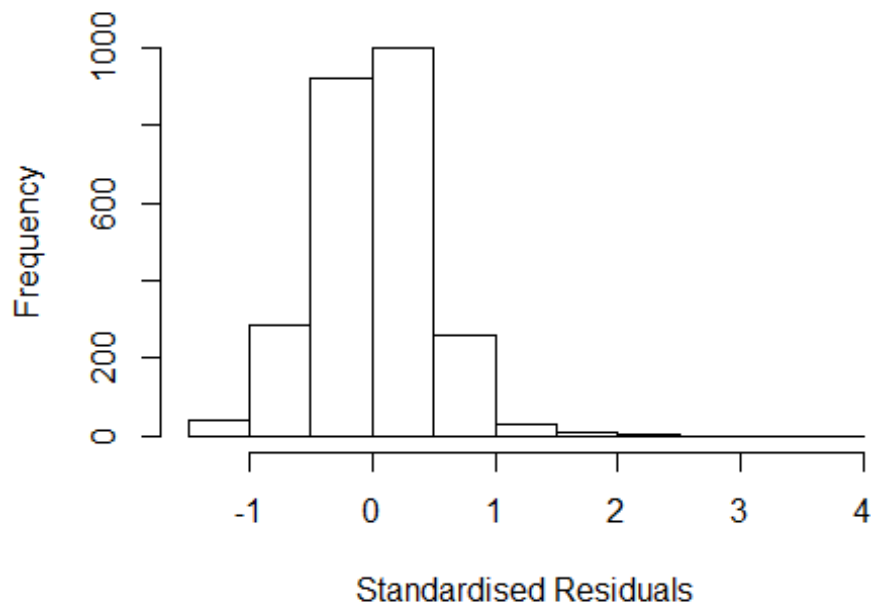
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 21, df = 1, p-value = 4e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.22777419415914"
## [1] "Male first author team size 2018 geometric mean: 3.04136908619496"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.71162903444405"
## [1] "Male last author team size 2018 geometric mean: 3.35365579553894"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 880, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.098 1      1.048
## LastAuthorFemale  1.070 1      1.034
## UniqueAuthors    1.164 4      1.019
## Year              1.274 16     1.008
```



## Residuals from first and last author and team size



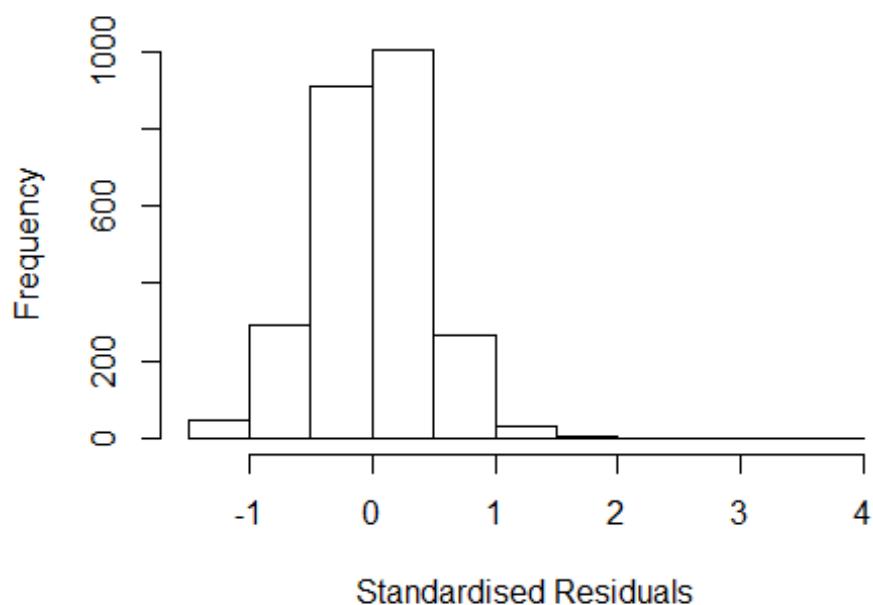
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2541 13244281317 4.541 2004      1315      2      3.624
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27794 -0.28088  0.00914  0.28651  3.62371
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18025    0.05325   22.16 < 2e-16 ***
## FirstAuthorFemale1 -0.04941    0.01860   -2.66  0.0080 **
## LastAuthorFemale1  0.02430    0.02381    1.02  0.3076
## UniqueAuthors2   -0.00678    0.04526   -0.15  0.8810
## UniqueAuthors3    0.02364    0.04383    0.54  0.5897
## UniqueAuthors4    0.03634    0.04406    0.82  0.4096
## UniqueAuthors5    0.09325    0.04241    2.20  0.0280 *
## Year1997          0.05384    0.05459    0.99  0.3241
## Year1998         -0.03784    0.04979   -0.76  0.4474
## Year1999         -0.10129    0.05029   -2.01  0.0441 *
```

```

## Year2000      -0.09981    0.05416   -1.84    0.0655 .
## Year2001      -0.04919    0.05030   -0.98    0.3283
## Year2002      -0.15235    0.05006   -3.04    0.0024 **
## Year2003      -0.22691    0.05340   -4.25    2.2e-05 ***
## Year2004      -0.25619    0.04952   -5.17    2.5e-07 ***
## Year2005      -0.25926    0.05130   -5.05    4.6e-07 ***
## Year2006      -0.21907    0.05108   -4.29    1.9e-05 ***
## Year2007      -0.12006    0.05755   -2.09    0.0371 *
## Year2008      -0.16640    0.05486   -3.03    0.0024 **
## Year2009      -0.09299    0.04838   -1.92    0.0547 .
## Year2010      -0.05418    0.05262   -1.03    0.3032
## Year2011      -0.11107    0.05343   -2.08    0.0377 *
## Year2012      -0.09411    0.06091   -1.54    0.1225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0491, Adjusted R-squared:  0.0408
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(1150,1240,1266,1693)
## are outliers with |weight| = 0 ( < 3.9e-05);
## 233 weights are ~ = 1. The remaining 2317 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0009 0.8660 0.9500 0.8990 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.92e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.082 1 1.040
## LastAuthorFemale 1.046 1 1.023
## Year 1.129 16 1.004

```

## Residuals from first and last author

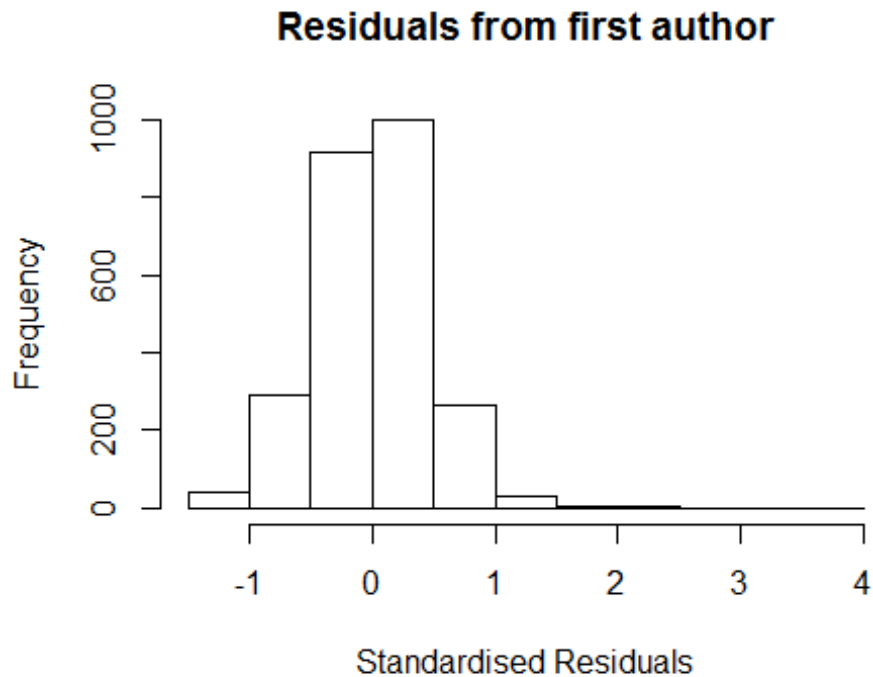


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2541 13244281317 4.541 2004      1315      2      3.575
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2869 -0.2781  0.0106  0.2855  3.5753
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2136     0.0389   31.18 < 2e-16 ***
## FirstAuthorFemale1 -0.0426     0.0186   -2.29  0.0223 *
## LastAuthorFemale1  0.0153     0.0236    0.65  0.5169
## Year1997          0.0580     0.0550    1.05  0.2917
## Year1998         -0.0358     0.0500   -0.72  0.4739
## Year1999         -0.0971     0.0502   -1.94  0.0531 .
## Year2000         -0.0972     0.0543   -1.79  0.0735 .
## Year2001         -0.0434     0.0507   -0.86  0.3921
## Year2002         -0.1403     0.0504   -2.78  0.0054 **
## Year2003         -0.2270     0.0539   -4.21 2.6e-05 ***
## Year2004         -0.2479     0.0502   -4.94 8.4e-07 ***
## Year2005         -0.2502     0.0520   -4.81 1.6e-06 ***
```

```

## Year2006          -0.2089      0.0519   -4.03  5.8e-05 ***
## Year2007          -0.1083      0.0573   -1.89  0.0589 .
## Year2008          -0.1455      0.0545   -2.67  0.0077 **
## Year2009          -0.0808      0.0486   -1.66  0.0967 .
## Year2010          -0.0438      0.0526   -0.83  0.4053
## Year2011          -0.0976      0.0533   -1.83  0.0674 .
## Year2012          -0.0840      0.0616   -1.36  0.1726
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.424
## Multiple R-squared:  0.0408, Adjusted R-squared:  0.034
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(1150,1240,1266,1693)
## are outliers with |weight| = 0 ( < 3.9e-05);
## 226 weights are ~ = 1. The remaining 2324 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0069 0.8650 0.9510 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.92e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.082 1          1.040
## Year              1.082 16          1.002

```

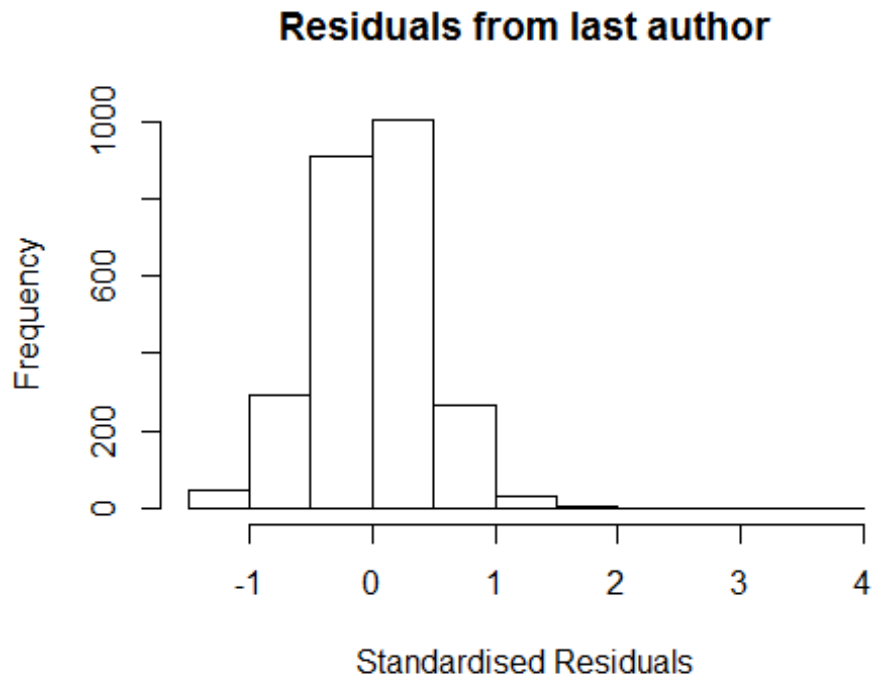


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2541 13244281317 4.541 2004      1315      2      3.575
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27334 -0.27651  0.00949  0.28703  3.57347
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2159    0.0387   31.39 < 2e-16 ***
## FirstAuthorFemale1 -0.0421    0.0186   -2.26  0.0240 *
## Year1997          0.0575    0.0550    1.05  0.2957
## Year1998         -0.0357    0.0500   -0.72  0.4746
## Year1999         -0.0964    0.0502   -1.92  0.0549 .
## Year2000         -0.0968    0.0543   -1.78  0.0747 .
## Year2001         -0.0434    0.0508   -0.85  0.3934
## Year2002         -0.1409    0.0504   -2.79  0.0052 **
## Year2003         -0.2266    0.0539   -4.20 2.8e-05 ***
## Year2004         -0.2483    0.0502   -4.94 8.1e-07 ***
## Year2005         -0.2509    0.0520   -4.83 1.5e-06 ***
## Year2006         -0.2092    0.0519   -4.03 5.8e-05 ***
```

```

## Year2007          -0.1070      0.0572   -1.87   0.0617 .
## Year2008          -0.1449      0.0546   -2.66   0.0080 **
## Year2009          -0.0804      0.0486   -1.65   0.0985 .
## Year2010          -0.0427      0.0526   -0.81   0.4168
## Year2011          -0.0980      0.0533   -1.84   0.0664 .
## Year2012          -0.0826      0.0615   -1.34   0.1791
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.424
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.0342
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(1150,1240,1266,1693)
## are outliers with |weight| = 0 ( < 3.9e-05);
## 230 weights are ~ = 1. The remaining 2320 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0074 0.8650 0.9510 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.92e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.045 1          1.022
## Year            1.045 16          1.001

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2541 13244281317 4.541 2004      1315      2      3.575
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2759 -0.2796  0.0111  0.2880  3.5882
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2019     0.0386   31.16 < 2e-16 ***
## LastAuthorFemale1  0.0130     0.0236    0.55  0.5815
## Year1997          0.0609     0.0549    1.11  0.2668
## Year1998         -0.0369     0.0501   -0.74  0.4611
## Year1999         -0.1011     0.0503   -2.01  0.0445 *
## Year2000         -0.0955     0.0544   -1.76  0.0792 .
## Year2001         -0.0411     0.0508   -0.81  0.4190
## Year2002         -0.1394     0.0505   -2.76  0.0058 **
## Year2003         -0.2320     0.0537   -4.32  1.6e-05 ***
## Year2004         -0.2491     0.0501   -4.98  6.9e-07 ***
## Year2005         -0.2519     0.0519   -4.85  1.3e-06 ***
## Year2006         -0.2132     0.0517   -4.13  3.8e-05 ***
```

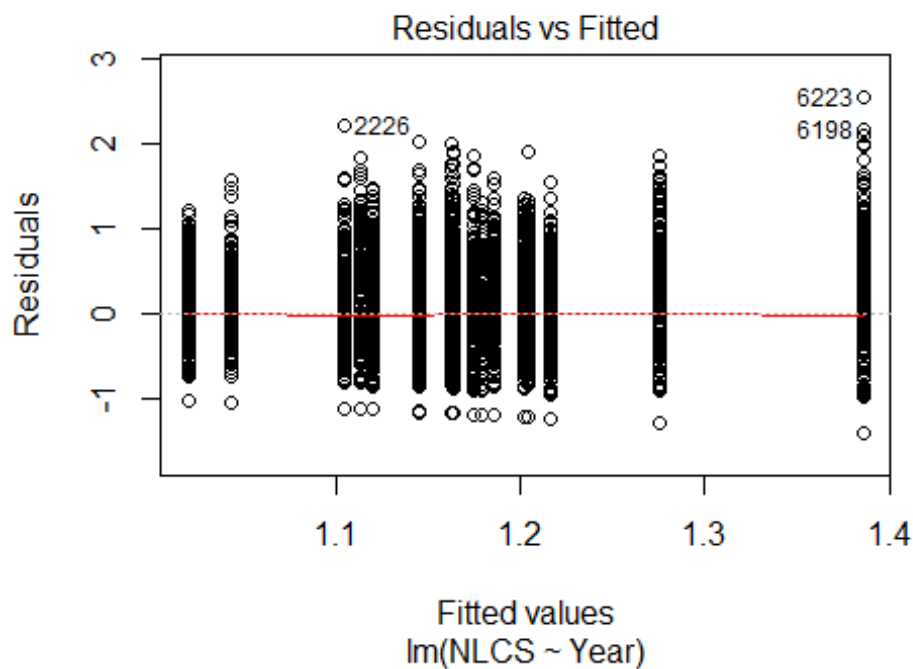
```

## Year2007          -0.1108      0.0573    -1.93    0.0533 .
## Year2008          -0.1496      0.0546    -2.74    0.0062 **
## Year2009          -0.0845      0.0486    -1.74    0.0824 .
## Year2010          -0.0468      0.0527    -0.89    0.3748
## Year2011          -0.0995      0.0532    -1.87    0.0618 .
## Year2012          -0.0903      0.0611    -1.48    0.1398
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0391, Adjusted R-squared:  0.0327
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 observations c(1150,1240,1266,1693)
## are outliers with |weight| = 0 ( < 3.9e-05);
## 225 weights are ~ = 1. The remaining 2325 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.005  0.867   0.951   0.900   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2554"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1400"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   306  277  310  264  380  315  251  239  269  285  307  397  358  369  379
## 2011 2012
##   359  327
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

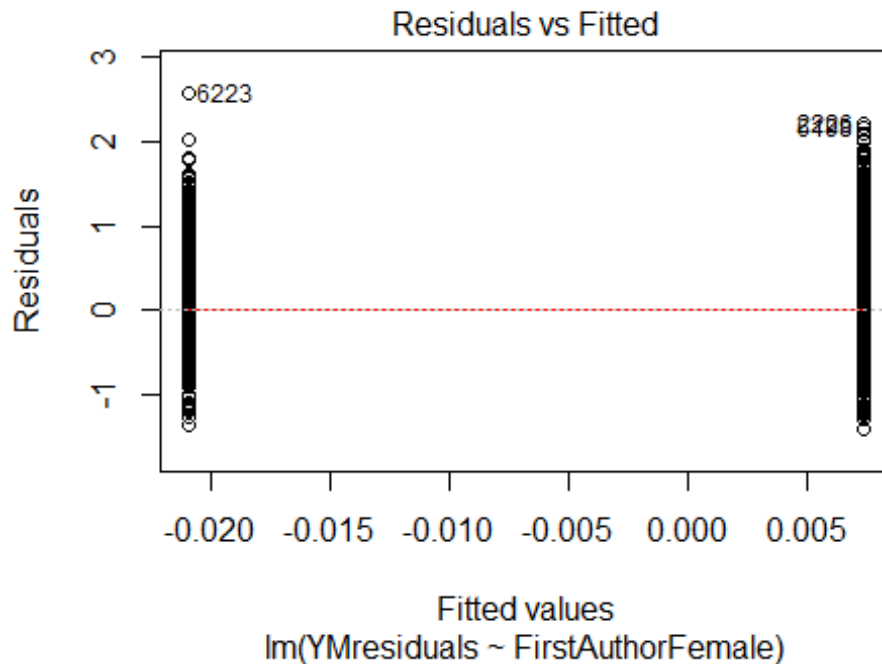
```



```
## 243 214 227 201 288 256 199 196 218 236 256 330 295 310 323
## 2011 2012
## 306 275
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 234 208 218 191 270 243 189 188 208 218 242 306 280 285 304
## 2011 2012
## 294 254
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 95, df = 16, p-value = 3e-13
```

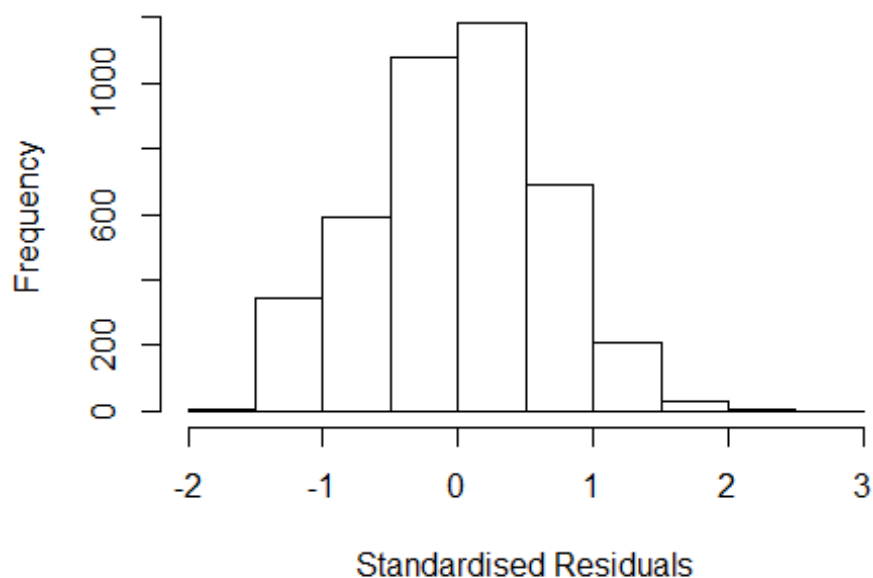


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.014, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 1.71012274412324"
## [1] "Male first author team size 2018 geometric mean: 1.66934461872387"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8900, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.59972017071299"
## [1] "Male last author team size 2018 geometric mean: 1.72757287704164"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7800, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.340 1      1.157
## LastAuthorFemale  1.347 1      1.161
## UniqueAuthors     1.104 4      1.012
## Year              1.124 16     1.004
```

## Residuals from first and last author and team size



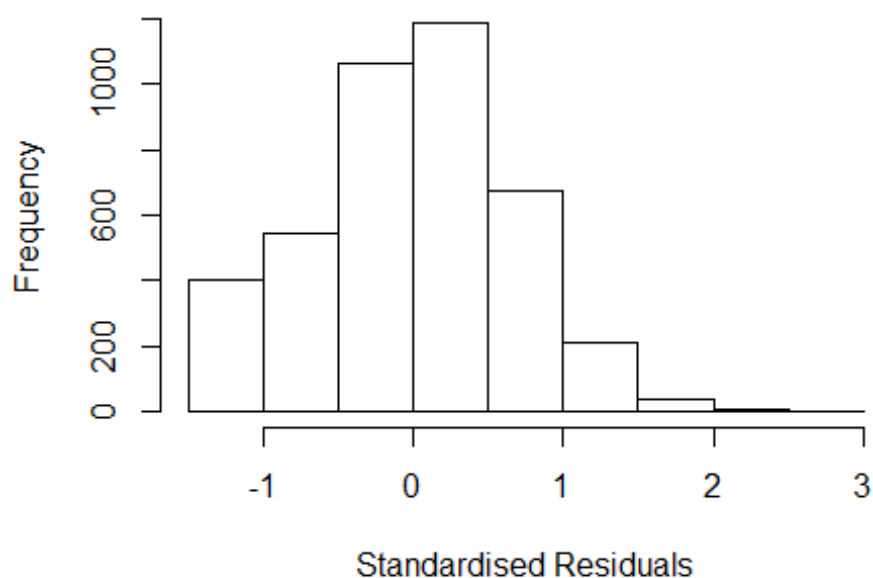
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6223 84858376384 3.929 2012    1400    1    2.74
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5538 -0.4413  0.0227  0.4463  2.7404
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01051    0.04726   21.38 < 2e-16 ***
## FirstAuthorFemale1 -0.04217    0.02715   -1.55  0.1204
## LastAuthorFemale1  0.00628    0.02743    0.23  0.8189
## UniqueAuthors2    0.20592    0.02428    8.48 < 2e-16 ***
## UniqueAuthors3    0.32305    0.03211   10.06 < 2e-16 ***
## UniqueAuthors4    0.32482    0.04926    6.59 4.8e-11 ***
## UniqueAuthors5    0.18750    0.07981    2.35  0.0189 *
## Year1997          0.09770    0.06380    1.53  0.1258
## Year1998          0.02515    0.06543    0.38  0.7007
## Year1999          0.06008    0.06197    0.97  0.3323
```

```

## Year2000          0.10651      0.05857      1.82      0.0691 .
## Year2001          -0.01933      0.06453     -0.30      0.7645
## Year2002          -0.03543      0.06621     -0.54      0.5926
## Year2003          -0.09807      0.06233     -1.57      0.1157
## Year2004           0.02593      0.06698      0.39      0.6987
## Year2005           0.04331      0.06654      0.65      0.5151
## Year2006          -0.11388      0.06016     -1.89      0.0584 .
## Year2007          -0.00102      0.06134     -0.02      0.9868
## Year2008           0.09057      0.06537      1.39      0.1660
## Year2009           0.03719      0.06409      0.58      0.5618
## Year2010           0.07096      0.06073      1.17      0.2427
## Year2011           0.15194      0.06730      2.26      0.0240 *
## Year2012           0.21397      0.07057      3.03      0.0024 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.652
## Multiple R-squared:  0.051, Adjusted R-squared:  0.0459
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 338 weights are ~= 1. The remaining 3794 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.038  0.867  0.949   0.909   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.338 1      1.157
## LastAuthorFemale  1.334 1      1.155
## Year              1.041 16      1.001

```

## Residuals from first and last author



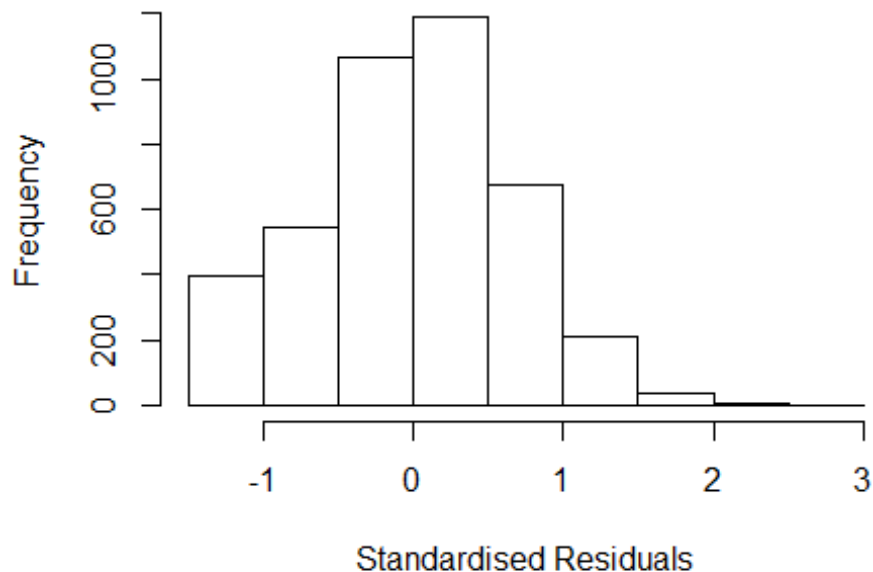
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6223 84858376384 3.929 2012    1400      1    2.604
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3611 -0.4454  0.0239  0.4435  2.6041
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09144    0.04898   22.29 < 2e-16 ***
## FirstAuthorFemale1 -0.03621    0.02796   -1.29  0.19542
## LastAuthorFemale1  0.00315    0.02809    0.11  0.91067
## Year1997         0.11592    0.06524    1.78  0.07568 .
## Year1998         0.03968    0.06682    0.59  0.55259
## Year1999         0.08698    0.06350    1.37  0.17080
## Year2000         0.14125    0.06007    2.35  0.01875 *
## Year2001         0.01079    0.06583    0.16  0.86985
## Year2002         0.00102    0.06763    0.02  0.98796
## Year2003        -0.05394    0.06419   -0.84  0.40081
## Year2004         0.06455    0.06989    0.92  0.35570
## Year2005         0.09162    0.06847    1.34  0.18090
```

```

## Year2006      -0.08157    0.06252   -1.30  0.19210
## Year2007      0.04279    0.06297    0.68  0.49683
## Year2008      0.13284    0.06671    1.99  0.04652 *
## Year2009      0.06526    0.06658    0.98  0.32706
## Year2010      0.10953    0.06286    1.74  0.08152 .
## Year2011      0.18394    0.06930    2.65  0.00798 **
## Year2012      0.26649    0.07289    3.66  0.00026 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.0113
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 348 weights are ~= 1. The remaining 3784 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.087  0.868  0.949  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1      1.013
## Year      1.026 16      1.001

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6223 84858376384 3.929 2012    1400      1    2.604
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3584 -0.4458  0.0228  0.4437  2.6051
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09169    0.04897   22.29 < 2e-16 ***
## FirstAuthorFemale1 -0.03447    0.02455   -1.40  0.16031
## Year1997        0.11601    0.06525    1.78  0.07549 .
## Year1998        0.03966    0.06683    0.59  0.55288
## Year1999        0.08706    0.06347    1.37  0.17024
## Year2000        0.14138    0.06006    2.35  0.01862 *
## Year2001        0.01086    0.06582    0.17  0.86893
## Year2002        0.00112    0.06762    0.02  0.98680
## Year2003       -0.05378    0.06414   -0.84  0.40176
## Year2004        0.06482    0.06978    0.93  0.35295
## Year2005        0.09166    0.06846    1.34  0.18068
## Year2006       -0.08141    0.06248   -1.30  0.19265
```

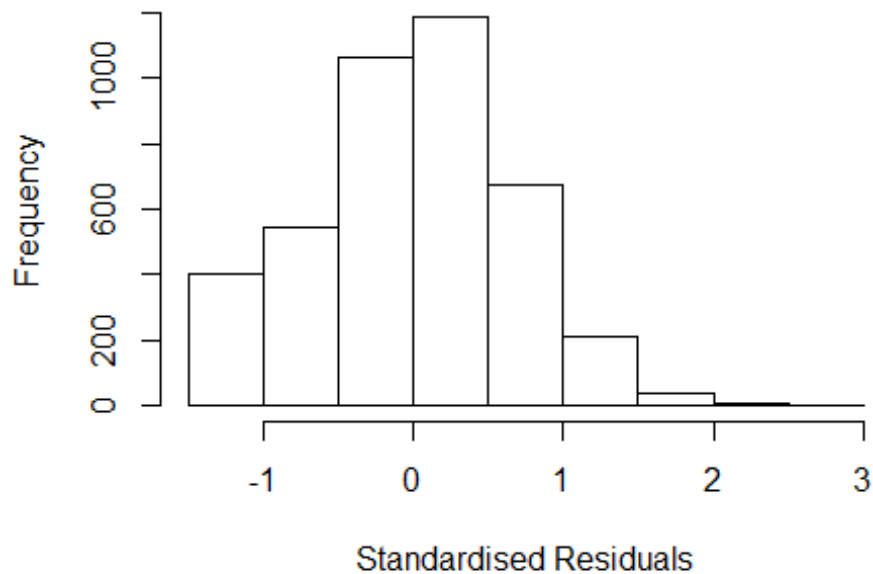
```

## Year2007          0.04287    0.06295    0.68  0.49585
## Year2008          0.13303    0.06665    2.00  0.04600 *
## Year2009          0.06536    0.06656    0.98  0.32620
## Year2010          0.10971    0.06284    1.75  0.08091 .
## Year2011          0.18412    0.06927    2.66  0.00789 **
## Year2012          0.26667    0.07278    3.66  0.00025 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0156, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 347 weights are ~= 1. The remaining 3785 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0864 0.8680 0.9490 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.022 1          1.011
## Year            1.022 16          1.001

```



## Residuals from last author



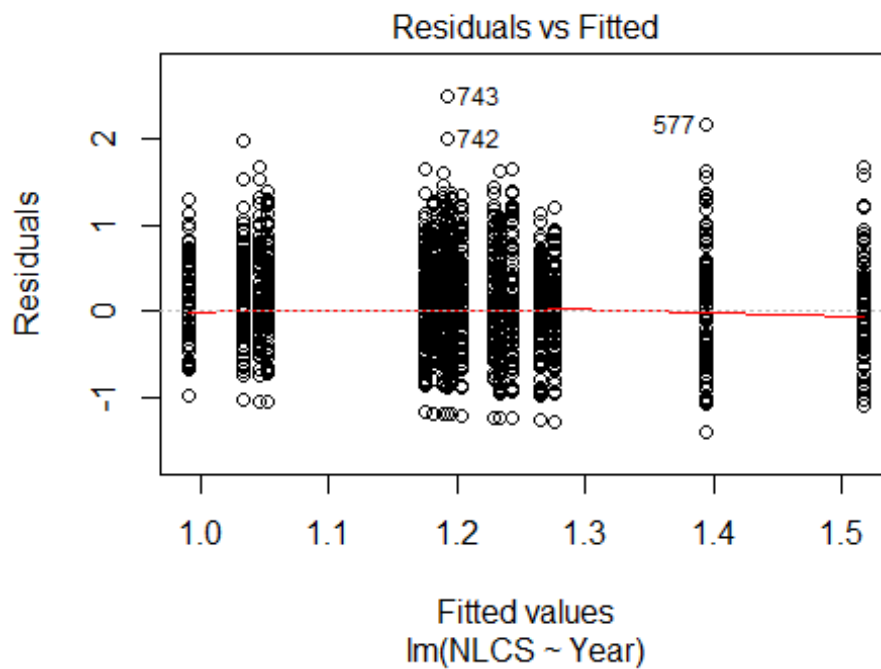
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6223 84858376384 3.929 2012    1400    1    2.604
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3539 -0.4472  0.0194  0.4455  2.5919
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.089276   0.048835   22.31 < 2e-16 ***
## LastAuthorFemale1 -0.016774   0.024730   -0.68  0.49763
## Year1997         0.116055   0.065228    1.78  0.07528 .
## Year1998         0.037297   0.066754    0.56  0.57638
## Year1999         0.084200   0.063422    1.33  0.18438
## Year2000         0.139347   0.060048    2.32  0.02036 *
## Year2001         0.009185   0.065769    0.14  0.88893
## Year2002        -0.000471   0.067650   -0.01  0.99444
## Year2003        -0.056815   0.064116   -0.89  0.37560
## Year2004         0.064061   0.069865    0.92  0.35924
## Year2005         0.088574   0.068213    1.30  0.19419
## Year2006        -0.084410   0.062407   -1.35  0.17627
```

```

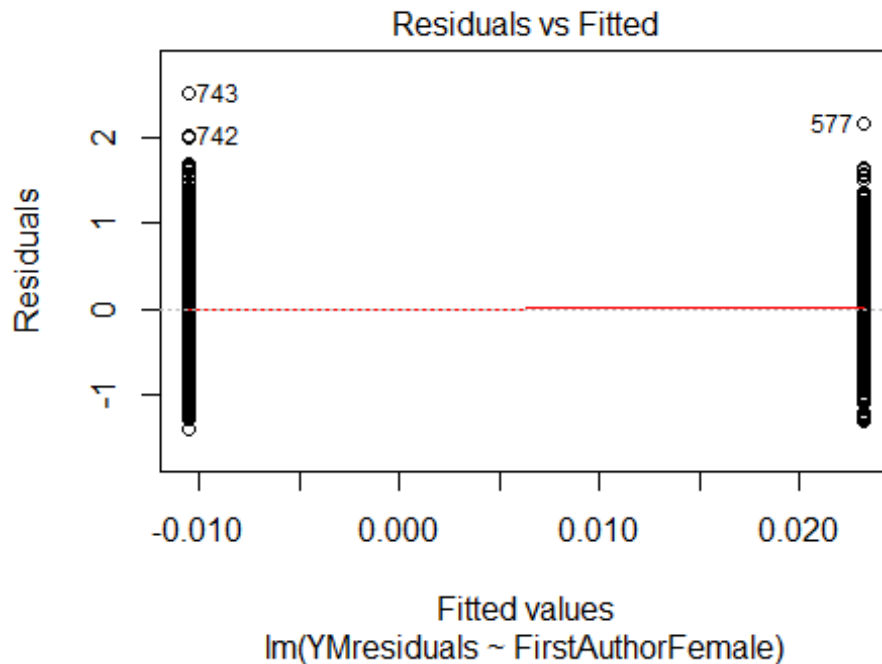
## Year2007      0.040846  0.062887  0.65  0.51604
## Year2008      0.129888  0.066633  1.95  0.05133 .
## Year2009      0.061519  0.066400  0.93  0.35425
## Year2010      0.106998  0.062760  1.70  0.08829 .
## Year2011      0.181287  0.069271  2.62  0.00890 **
## Year2012      0.264643  0.072936  3.63  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0153, Adjusted R-squared:  0.0112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 344 weights are ~= 1. The remaining 3788 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0911 0.8680 0.9500 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4132"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1401"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 146 137 177 162 186 128 129 135 170 126 148 183 186 182 186
## 2011 2012
## 150 168
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 117 114 113 108 128 102 112 120 135 106 116 160 154 149 161
## 2011 2012

```

```
## 127 133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 107 112 102 126 98 100 117 127 98 108 153 140 136 146
## 2011 2012
## 112 122
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 37, df = 16, p-value = 0.002
```

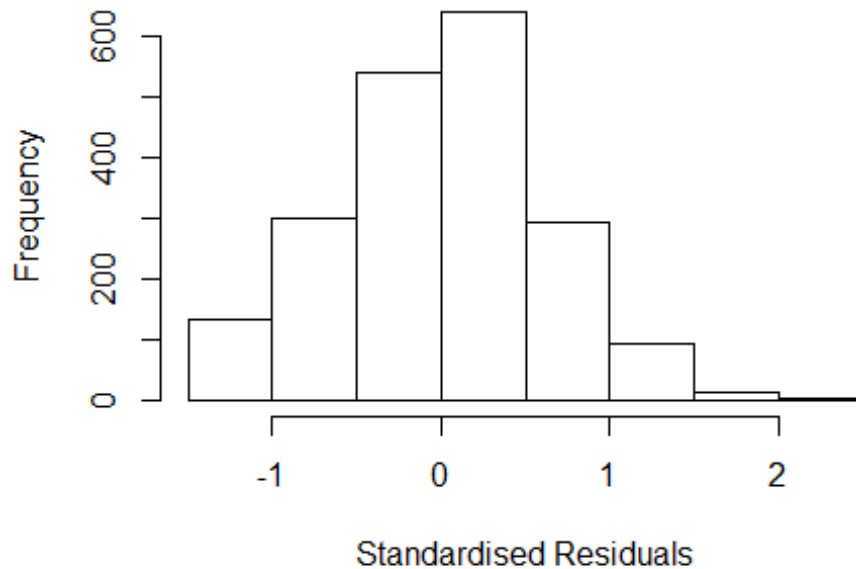


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0098, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 1.9484221923315"
## [1] "Male first author team size 2018 geometric mean: 1.72405175603237"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.83959213767928"
## [1] "Male last author team size 2018 geometric mean: 1.7842681226699"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.480 1          1.216
## LastAuthorFemale  1.469 1          1.212
## UniqueAuthors    1.180 4          1.021
## Year              1.285 16         1.008
```

## Residuals from first and last author and team size



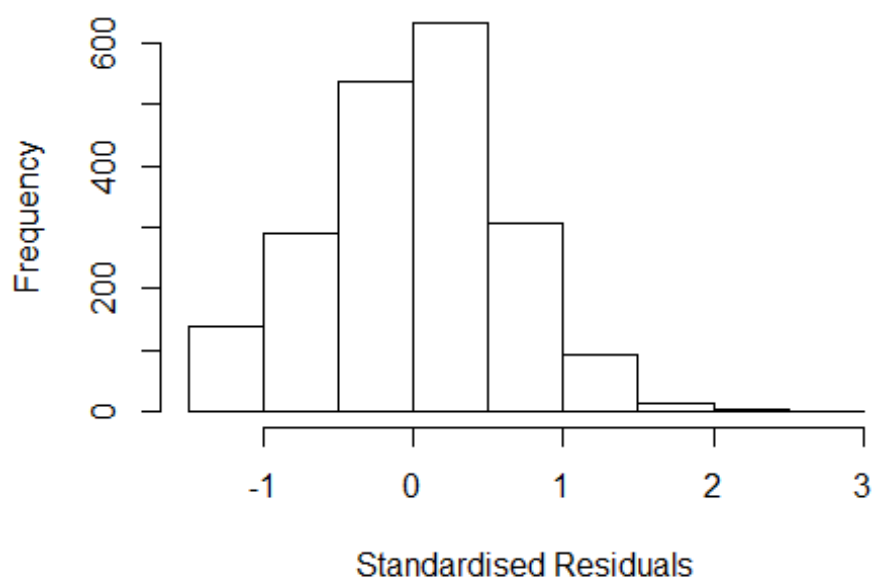
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4725 -0.4167 0.0254 0.4093 2.4318
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8834 0.0583 15.16 < 2e-16 ***
## FirstAuthorFemale1 0.0148 0.0363 0.41 0.68314
## LastAuthorFemale1 0.0208 0.0358 0.58 0.56169
## UniqueAuthors2 0.2147 0.0323 6.65 3.8e-11 ***
## UniqueAuthors3 0.2251 0.0392 5.74 1.1e-08 ***
## UniqueAuthors4 0.3171 0.0810 3.92 9.3e-05 ***
## UniqueAuthors5 0.3146 0.1047 3.00 0.00270 **
## Year1997 0.0357 0.0852 0.42 0.67516
## Year1998 0.0272 0.0850 0.32 0.74891
## Year1999 0.3744 0.0967 3.87 0.00011 ***
```

```

## Year2000          0.1541      0.0829      1.86  0.06325 .
## Year2001          0.1888      0.0932      2.03  0.04281 *
## Year2002          0.2551      0.0786      3.24  0.00120 **
## Year2003          0.2829      0.0748      3.78  0.00016 ***
## Year2004          0.1971      0.0774      2.54  0.01101 *
## Year2005          0.4667      0.0821      5.68  1.5e-08 ***
## Year2006          0.1853      0.0832      2.23  0.02611 *
## Year2007          0.1444      0.0705      2.05  0.04068 *
## Year2008          0.2078      0.0792      2.62  0.00877 **
## Year2009          0.0644      0.0774      0.83  0.40607
## Year2010          0.1553      0.0748      2.08  0.03806 *
## Year2011          0.1957      0.0815      2.40  0.01642 *
## Year2012          0.2026      0.0794      2.55  0.01084 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0679, Adjusted R-squared:  0.0576
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 1844 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0678 0.8700 0.9490 0.9050 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.96e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.470 1          1.212
## LastAuthorFemale 1.459 1          1.208
## Year 1.107 16          1.003

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 743 84986127796 3.684 2000      1401      1      2.521
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3511 -0.4143  0.0249  0.4077  2.5214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9753    0.0581   16.78 < 2e-16 ***
## FirstAuthorFemale1 0.0242    0.0369    0.65  0.51296
## LastAuthorFemale1 0.0308    0.0363    0.85  0.39567
## Year1997        0.0230    0.0868    0.27  0.79077
## Year1998        0.0229    0.0885    0.26  0.79556
## Year1999        0.3758    0.0970    3.87  0.00011 ***
## Year2000        0.1872    0.0849    2.21  0.02750 *
## Year2001        0.2179    0.0966    2.26  0.02414 *
## Year2002        0.2802    0.0802    3.50  0.00048 ***
## Year2003        0.2999    0.0748    4.01  6.3e-05 ***
## Year2004        0.2299    0.0777    2.96  0.00314 **
## Year2005        0.5007    0.0818    6.12  1.1e-09 ***
```

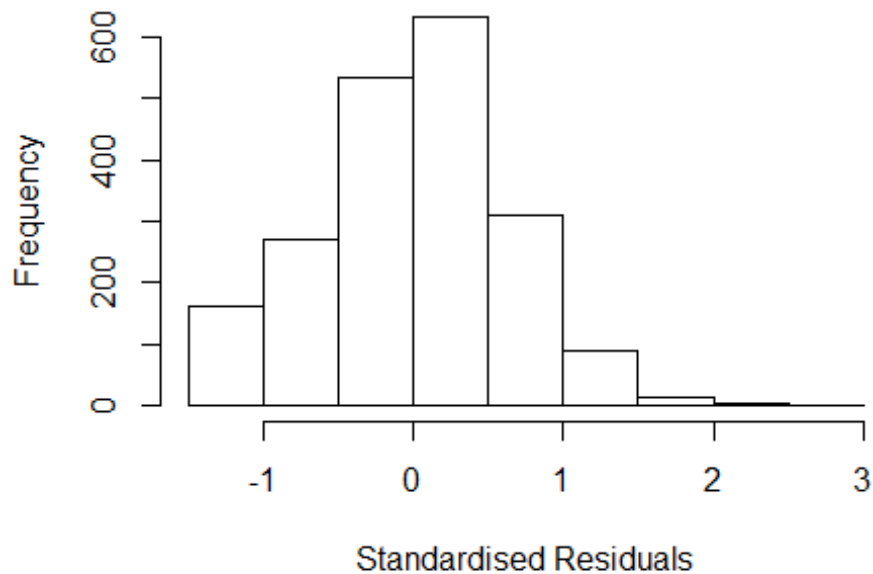
```

## Year2006          0.2150      0.0874      2.46  0.01396 *
## Year2007          0.1878      0.0712      2.64  0.00842 **
## Year2008          0.2078      0.0816      2.55  0.01091 *
## Year2009          0.0621      0.0803      0.77  0.43961
## Year2010          0.1735      0.0763      2.27  0.02313 *
## Year2011          0.1953      0.0829      2.36  0.01858 *
## Year2012          0.2268      0.0803      2.82  0.00481 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0362, Adjusted R-squared:  0.0275
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 197 weights are ~= 1. The remaining 1819 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0467 0.8630 0.9470 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1      1.023
## Year      1.047 16      1.001

```



## Residuals from first author



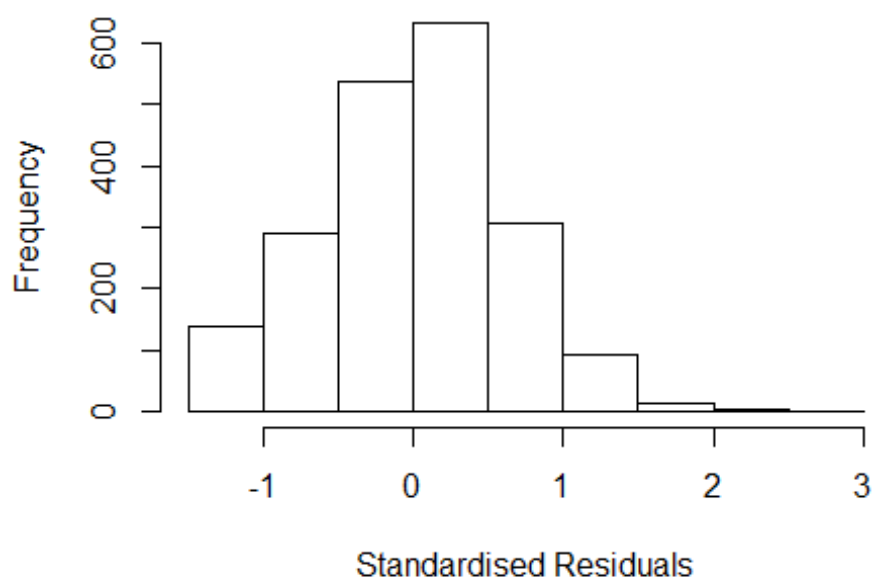
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 743 84986127796 3.684 2000      1401      1      2.521
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3553 -0.4155  0.0249  0.4045  2.5175
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9802    0.0577   17.00 < 2e-16 ***
## FirstAuthorFemale1 0.0413    0.0312    1.33  0.18489
## Year1997        0.0214    0.0868    0.25  0.80561
## Year1998        0.0238    0.0884    0.27  0.78817
## Year1999        0.3751    0.0970    3.87  0.00011 ***
## Year2000        0.1863    0.0847    2.20  0.02797 *
## Year2001        0.2207    0.0964    2.29  0.02219 *
## Year2002        0.2789    0.0802    3.48  0.00052 ***
## Year2003        0.2969    0.0746    3.98  7.2e-05 ***
## Year2004        0.2269    0.0776    2.93  0.00348 **
## Year2005        0.5011    0.0817    6.13  1.0e-09 ***
## Year2006        0.2140    0.0875    2.44  0.01458 *
```

```

## Year2007          0.1857      0.0711      2.61  0.00906 **
## Year2008          0.2075      0.0814      2.55  0.01093 *
## Year2009          0.0618      0.0804      0.77  0.44236
## Year2010          0.1749      0.0764      2.29  0.02226 *
## Year2011          0.1956      0.0828      2.36  0.01831 *
## Year2012          0.2262      0.0804      2.81  0.00496 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0358, Adjusted R-squared:  0.0276
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 193 weights are ~= 1. The remaining 1823 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0477 0.8610 0.9480 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.039 1      1.020
## Year      1.039 16      1.001

```

## Residuals from last author



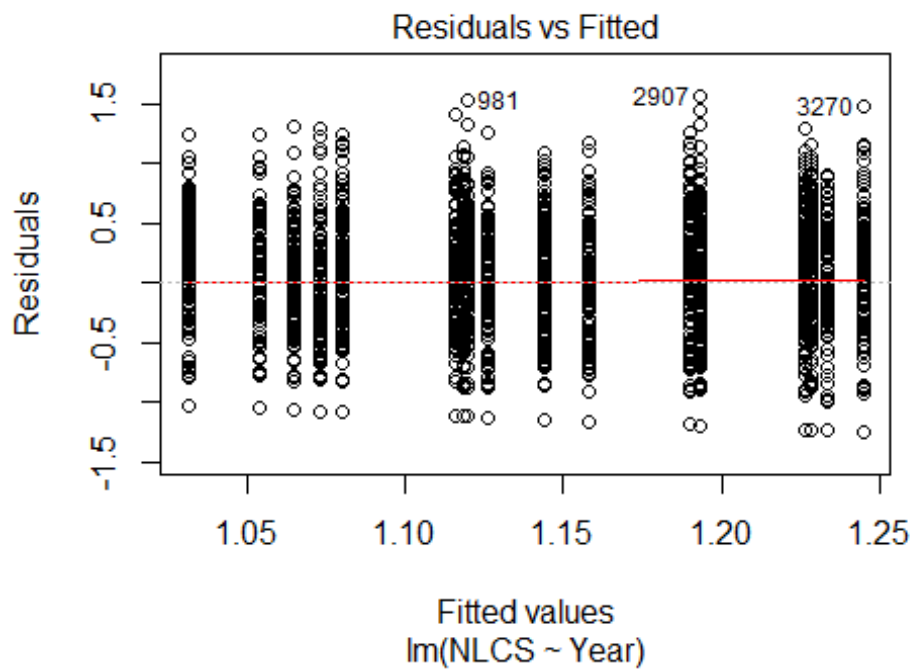
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 743 84986127796 3.684 2000      1401      1      2.521
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3542 -0.4139  0.0256  0.4050  2.5177
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9786    0.0578   16.93 < 2e-16 ***
## LastAuthorFemale1 0.0443    0.0306    1.45  0.14818
## Year1997        0.0243    0.0868    0.28  0.77982
## Year1998        0.0212    0.0883    0.24  0.81048
## Year1999        0.3756    0.0970    3.87  0.00011 ***
## Year2000        0.1877    0.0847    2.21  0.02689 *
## Year2001        0.2157    0.0963    2.24  0.02517 *
## Year2002        0.2800    0.0801    3.49  0.00049 ***
## Year2003        0.3008    0.0747    4.03  5.9e-05 ***
## Year2004        0.2306    0.0777    2.97  0.00304 **
## Year2005        0.5011    0.0818    6.12  1.1e-09 ***
## Year2006        0.2174    0.0872    2.49  0.01277 *
```

```

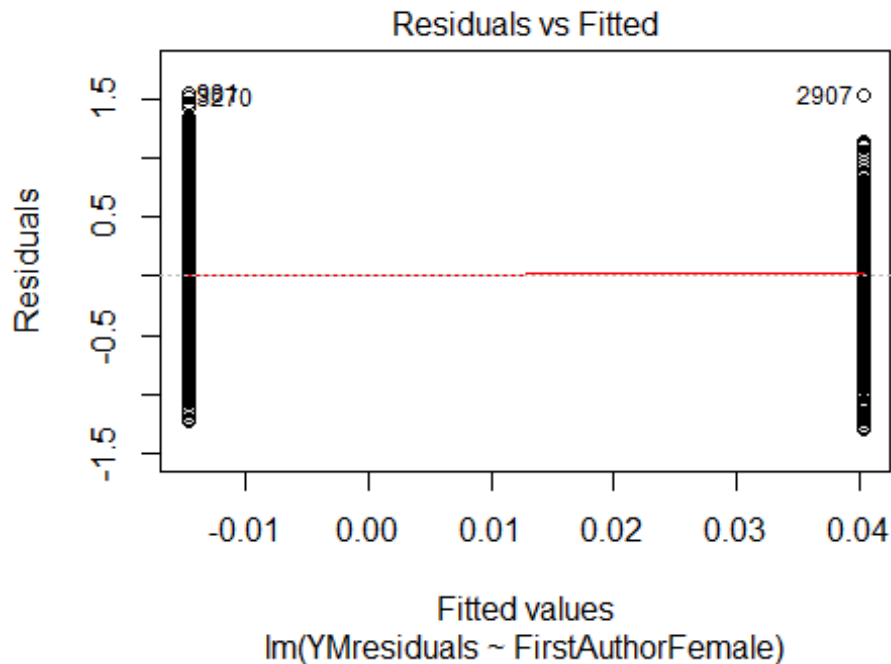
## Year2007          0.1892      0.0712      2.66  0.00796 **
## Year2008          0.2082      0.0816      2.55  0.01084 *
## Year2009          0.0615      0.0803      0.77  0.44362
## Year2010          0.1720      0.0761      2.26  0.02399 *
## Year2011          0.1946      0.0830      2.35  0.01909 *
## Year2012          0.2275      0.0804      2.83  0.00470 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.036, Adjusted R-squared:  0.0278
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 191 weights are ~= 1. The remaining 1825 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0481 0.8620 0.9480 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2016"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1402"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 139 139 139 155 182 150 149 93 160 144 193 209 200 220 239
## 2011 2012
## 181 205
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 114 113 116 124 150 115 125 82 132 121 167 169 166 179 194
## 2011 2012

```

```
## 158 171
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 111 110 113 120 138 108 120 74 123 114 153 156 154 167 183
## 2011 2012
## 145 155
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, df = 16, p-value = 0.2
```

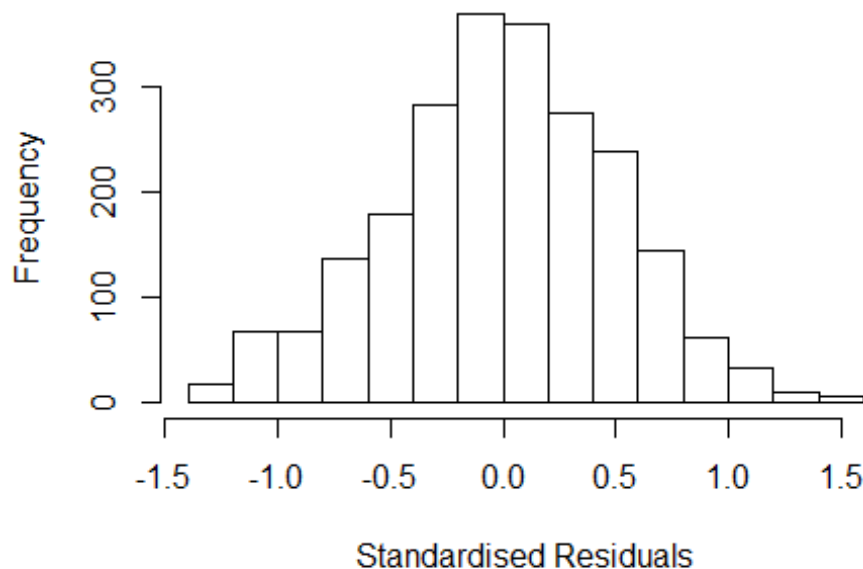


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.9, df = 1, p-value = 0.009
```



```
## [1] "Female first author team size 2018 geometric mean: 1.66347933893548"
## [1] "Male first author team size 2018 geometric mean: 1.63251220296247"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.45311183923054"
## [1] "Male last author team size 2018 geometric mean: 1.71677410000075"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.451 1 1.205
## LastAuthorFemale 1.448 1 1.203
## UniqueAuthors 1.192 4 1.022
## Year 1.250 16 1.007
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27354 -0.32773  0.00394  0.33850  1.54026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.014088   0.049958  20.30 < 2e-16 ***
## FirstAuthorFemale1 0.028755   0.028133   1.02  0.30684
## LastAuthorFemale1 0.058071   0.029141   1.99  0.04641 *
## UniqueAuthors2    0.061751   0.025761   2.40  0.01661 *
## UniqueAuthors3    0.083553   0.031816   2.63  0.00870 **
## UniqueAuthors4    0.167878   0.053951   3.11  0.00188 **
## UniqueAuthors5    0.429603   0.124285   3.46  0.00056 ***
## Year1997          -0.000172   0.067081   0.00  0.99795
## Year1998           0.003773   0.067970   0.06  0.95574
## Year1999           0.052903   0.064947   0.81  0.41542
```

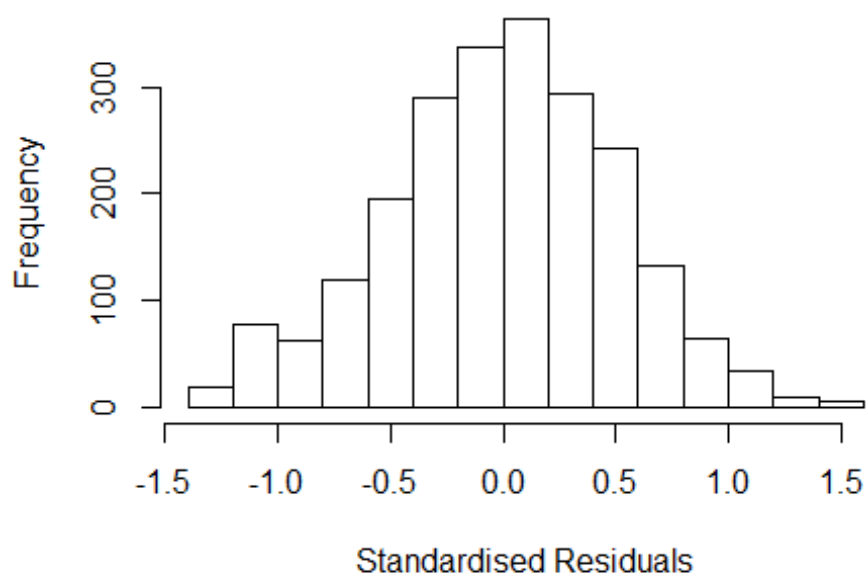
```

## Year2000      -0.004328    0.065034   -0.07   0.94695
## Year2001      0.045984    0.067837    0.68   0.49793
## Year2002      0.072723    0.063327    1.15   0.25094
## Year2003      0.201446    0.074062    2.72   0.00658 **
## Year2004     -0.016731    0.072050   -0.23   0.81640
## Year2005      0.031616    0.068753    0.46   0.64566
## Year2006      0.047769    0.064645    0.74   0.46003
## Year2007      0.083145    0.061746    1.35   0.17826
## Year2008      0.136401    0.068194    2.00   0.04560 *
## Year2009      0.137481    0.064283    2.14   0.03257 *
## Year2010      0.117829    0.062678    1.88   0.06025 .
## Year2011      0.156496    0.068188    2.30   0.02182 *
## Year2012      0.149143    0.070232    2.12   0.03382 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0329, Adjusted R-squared:  0.0233
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 190 weights are ~= 1. The remaining 2054 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.316  0.865   0.950   0.904   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.46e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.441 1          1.201
## LastAuthorFemale  1.436 1          1.198
## Year              1.078 16          1.002

```



## Residuals from first and last author



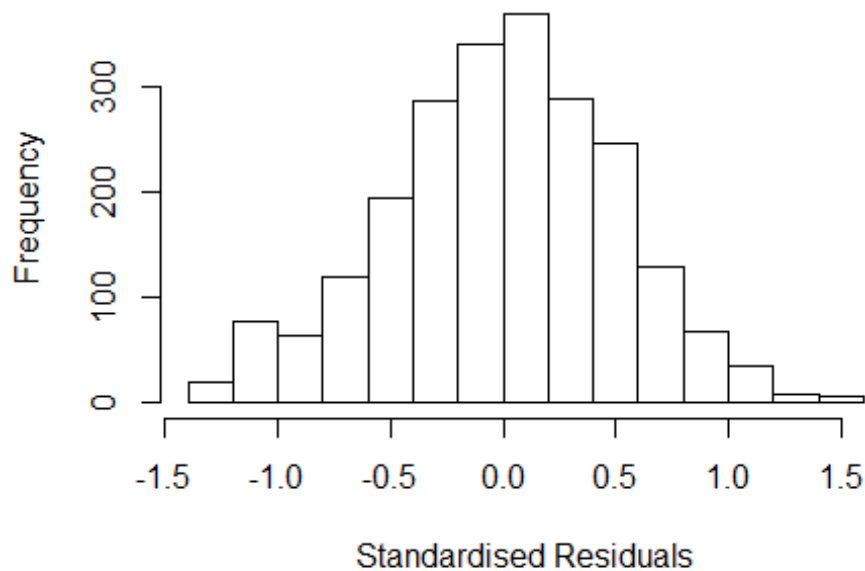
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3053 -0.3327 0.0114 0.3372 1.5376
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.041537 0.049117 21.21 <2e-16 ***
## FirstAuthorFemale1 0.035026 0.028196 1.24 0.2143
## LastAuthorFemale1 0.055309 0.029106 1.90 0.0575 .
## Year1997 -0.004852 0.067170 -0.07 0.9424
## Year1998 0.005259 0.067972 0.08 0.9383
## Year1999 0.059661 0.065082 0.92 0.3594
## Year2000 0.000971 0.065566 0.01 0.9882
## Year2001 0.068898 0.068555 1.01 0.3150
## Year2002 0.088583 0.063551 1.39 0.1635
## Year2003 0.212939 0.074814 2.85 0.0045 **
## Year2004 -0.010454 0.072340 -0.14 0.8851
## Year2005 0.043202 0.068449 0.63 0.5280
```

```

## Year2006          0.057773    0.065169    0.89    0.3754
## Year2007          0.107021    0.061666    1.74    0.0828 .
## Year2008          0.146839    0.067439    2.18    0.0296 *
## Year2009          0.152290    0.064117    2.38    0.0176 *
## Year2010          0.132524    0.062860    2.11    0.0351 *
## Year2011          0.173461    0.067888    2.56    0.0107 *
## Year2012          0.166905    0.070545    2.37    0.0181 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.0157
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 184 weights are ~= 1. The remaining 2060 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.871  0.950  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1         1.022
## Year              1.044 16         1.001

```

## Residuals from first author



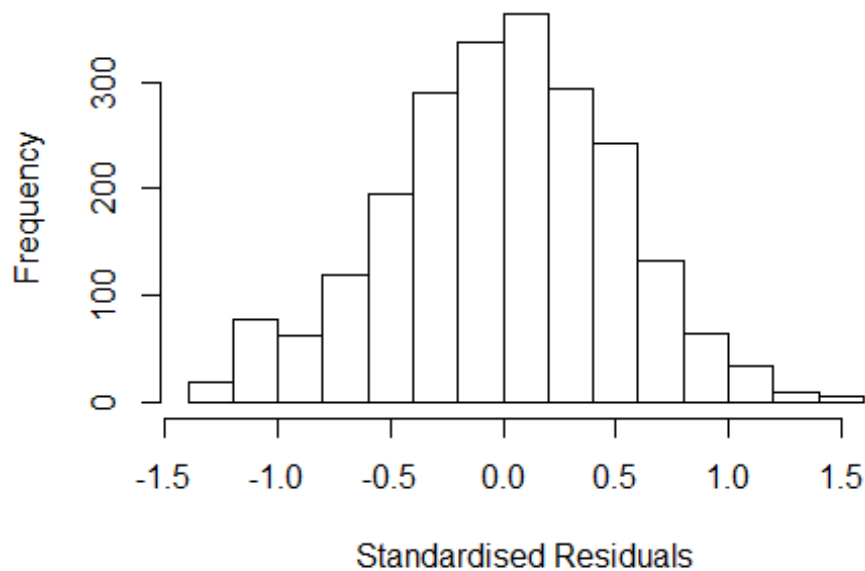
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2864 -0.3291 0.0105 0.3381 1.5322
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04491 0.04903 21.31 <2e-16 ***
## FirstAuthorFemale1 0.06397 0.02405 2.66 0.0079 **
## Year1997 -0.00465 0.06694 -0.07 0.9446
## Year1998 0.00651 0.06805 0.10 0.9238
## Year1999 0.06140 0.06526 0.94 0.3469
## Year2000 0.00212 0.06551 0.03 0.9742
## Year2001 0.07084 0.06846 1.03 0.3009
## Year2002 0.09082 0.06348 1.43 0.1526
## Year2003 0.21678 0.07484 2.90 0.0038 **
## Year2004 -0.01101 0.07238 -0.15 0.8791
## Year2005 0.04325 0.06826 0.63 0.5264
## Year2006 0.05941 0.06522 0.91 0.3625
```

```

## Year2007          0.11149    0.06164    1.81    0.0706 .
## Year2008          0.15015    0.06727    2.23    0.0257 *
## Year2009          0.15756    0.06385    2.47    0.0137 *
## Year2010          0.13367    0.06297    2.12    0.0339 *
## Year2011          0.17749    0.06767    2.62    0.0088 **
## Year2012          0.17066    0.07056    2.42    0.0157 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.0221, Adjusted R-squared:  0.0147
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 2066 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.324  0.871  0.950  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.042 1          1.021
## Year              1.042 16          1.001

```

## Residuals from last author



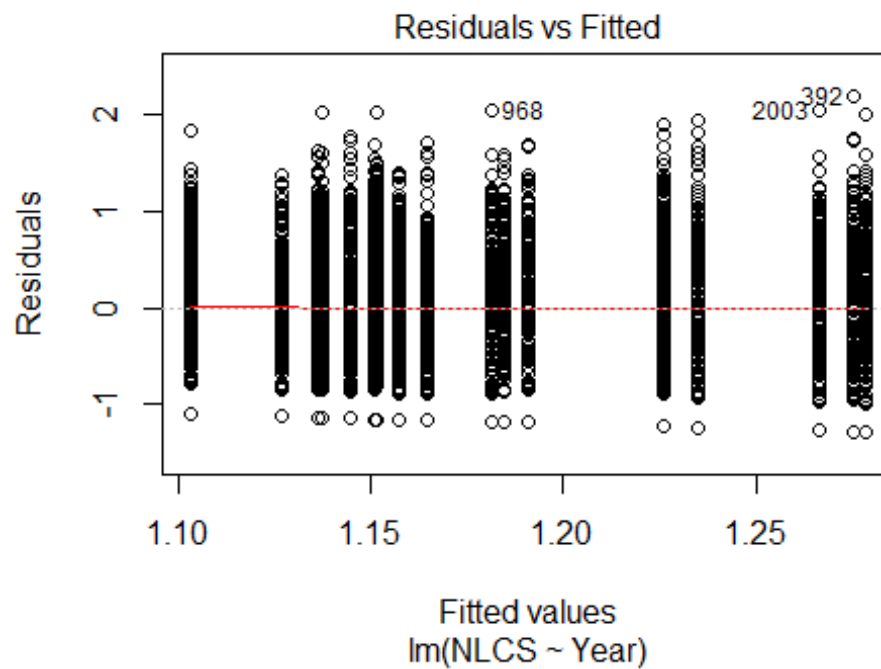
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2944 -0.3275  0.0103  0.3358  1.5347
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04495    0.04889   21.37  <2e-16 ***
## LastAuthorFemale1 0.07503    0.02481    3.02  0.0025 **
## Year1997       -0.00364    0.06724   -0.05  0.9569
## Year1998        0.00540    0.06788    0.08  0.9367
## Year1999        0.06219    0.06475    0.96  0.3369
## Year2000        0.00221    0.06556    0.03  0.9731
## Year2001        0.06836    0.06848    1.00  0.3183
## Year2002        0.08951    0.06348    1.41  0.1586
## Year2003        0.21488    0.07463    2.88  0.0040 **
## Year2004       -0.01012    0.07238   -0.14  0.8888
## Year2005        0.04618    0.06840    0.68  0.4997
## Year2006        0.06084    0.06490    0.94  0.3486
```

```

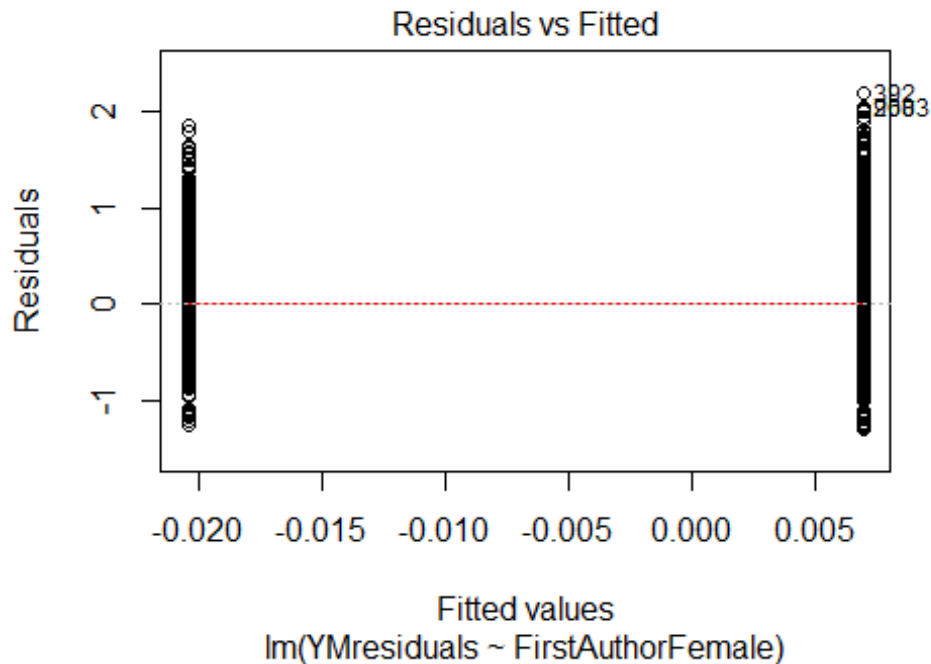
## Year2007      0.10550      0.06153      1.71      0.0865 .
## Year2008      0.14942      0.06720      2.22      0.0263 *
## Year2009      0.15415      0.06402      2.41      0.0161 *
## Year2010      0.13552      0.06269      2.16      0.0307 *
## Year2011      0.17445      0.06773      2.58      0.0101 *
## Year2012      0.17070      0.07013      2.43      0.0150 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.499
## Multiple R-squared:  0.023, Adjusted R-squared:  0.0155
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 188 weights are ~= 1. The remaining 2056 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.323  0.871  0.950  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2244"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1403"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 222 224 246 252 312 305 261 295 278 307 365 434 418 507 545
## 2011 2012
## 487 497
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 158 160 188 187 220 205 206 233 234 248 286 355 349 422 449
## 2011 2012

```

```
## 402 422
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 150 157 180 176 204 192 198 222 218 236 265 328 325 392 414
## 2011 2012
## 386 390
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 1e-04
```



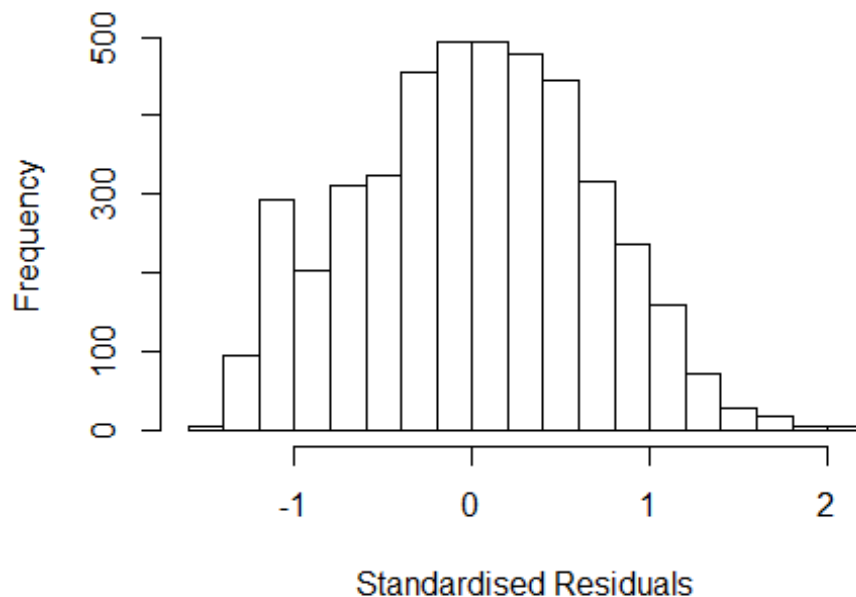
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 1.83293604771349"
## [1] "Male first author team size 2018 geometric mean: 1.68840161338848"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.84808806324931"
## [1] "Male last author team size 2018 geometric mean: 1.68407587301908"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.388 1          1.178
## LastAuthorFemale  1.374 1          1.172
## UniqueAuthors    1.066 4          1.008
## Year             1.079 16          1.002
```



## Residuals from first and last author and team size



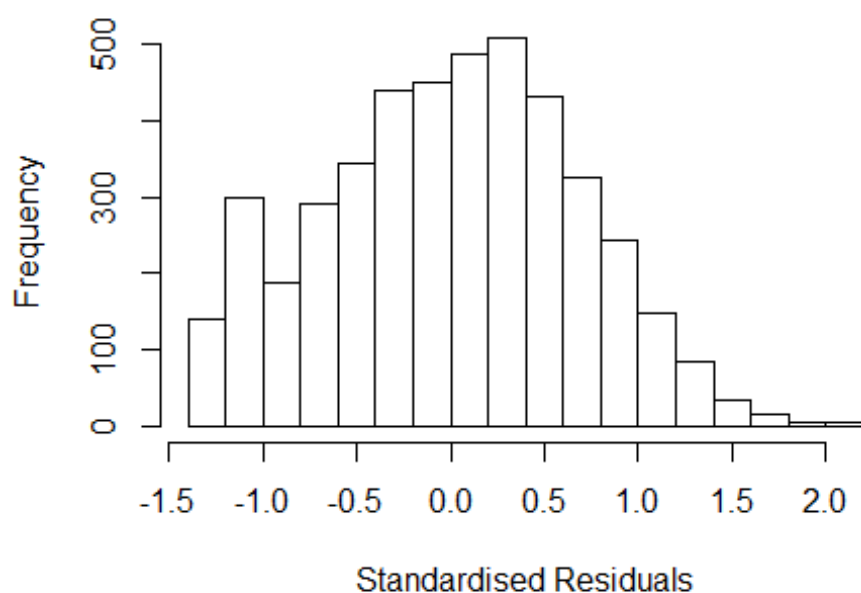
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5108 -0.4691  0.0187  0.4796  2.1422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12057    0.07560   14.82  < 2e-16 ***
## FirstAuthorFemale1 -0.03401    0.02764   -1.23    0.22
## LastAuthorFemale1 -0.00264    0.02846   -0.09    0.93
## UniqueAuthors2     0.18829    0.02383    7.90 3.5e-15 ***
## UniqueAuthors3     0.34022    0.03165   10.75 < 2e-16 ***
## UniqueAuthors4     0.28705    0.05844    4.91 9.3e-07 ***
## UniqueAuthors5     0.06433    0.09949    0.65    0.52
## Year1997           0.06576    0.10100    0.65    0.52
## Year1998           0.02488    0.09305    0.27    0.79
## Year1999          -0.00143    0.09251   -0.02    0.99
```

```

## Year2000      -0.02949    0.08840   -0.33    0.74
## Year2001      0.05003    0.09424    0.53    0.60
## Year2002      0.05723    0.09178    0.62    0.53
## Year2003     -0.09538    0.08573   -1.11    0.27
## Year2004     -0.09147    0.08777   -1.04    0.30
## Year2005     -0.07930    0.08543   -0.93    0.35
## Year2006     -0.07757    0.08400   -0.92    0.36
## Year2007     -0.10450    0.08365   -1.25    0.21
## Year2008     -0.08263    0.08337   -0.99    0.32
## Year2009     -0.07757    0.08280   -0.94    0.35
## Year2010     -0.12420    0.08182   -1.52    0.13
## Year2011     -0.01414    0.08328   -0.17    0.87
## Year2012     -0.08872    0.08286   -1.07    0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.0372, Adjusted R-squared:  0.0324
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 371 weights are ~= 1. The remaining 4062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.328  0.871  0.950  0.916  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.374 1      1.172
## LastAuthorFemale  1.369 1      1.170
## Year              1.031 16      1.001

```

## Residuals from first and last author



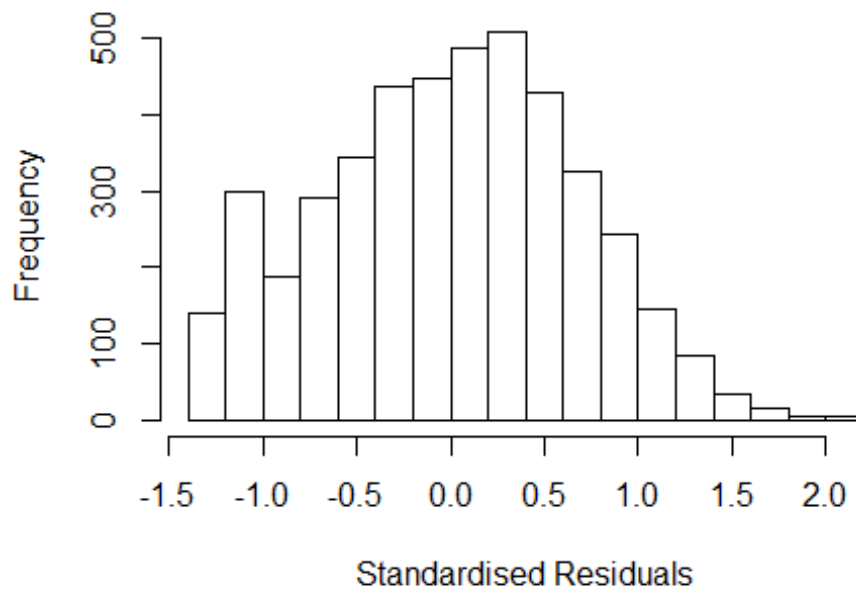
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2889 -0.4840 0.0303 0.4782 2.1967
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20347 0.07531 15.98 <2e-16 ***
## FirstAuthorFemale1 -0.01938 0.02805 -0.69 0.49
## LastAuthorFemale1 -0.00250 0.02896 -0.09 0.93
## Year1997 0.06487 0.10254 0.63 0.53
## Year1998 0.02802 0.09329 0.30 0.76
## Year1999 0.00341 0.09265 0.04 0.97
## Year2000 -0.01595 0.08859 -0.18 0.86
## Year2001 0.06898 0.09485 0.73 0.47
## Year2002 0.08548 0.09275 0.92 0.36
## Year2003 -0.08488 0.08654 -0.98 0.33
## Year2004 -0.07152 0.08950 -0.80 0.42
## Year2005 -0.04092 0.08577 -0.48 0.63
```

```

## Year2006      -0.05476    0.08498   -0.64    0.52
## Year2007      -0.06994    0.08396   -0.83    0.40
## Year2008      -0.05314    0.08372   -0.63    0.53
## Year2009      -0.05662    0.08302   -0.68    0.50
## Year2010      -0.09393    0.08206   -1.14    0.25
## Year2011       0.02084    0.08387    0.25    0.80
## Year2012      -0.05794    0.08343   -0.69    0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.00591,    Adjusted R-squared:  0.00186
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 354 weights are ~= 1. The remaining 4079 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.319  0.878  0.951  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.02 1      1.010
## Year              1.02 16      1.001

```

## Residuals from first author



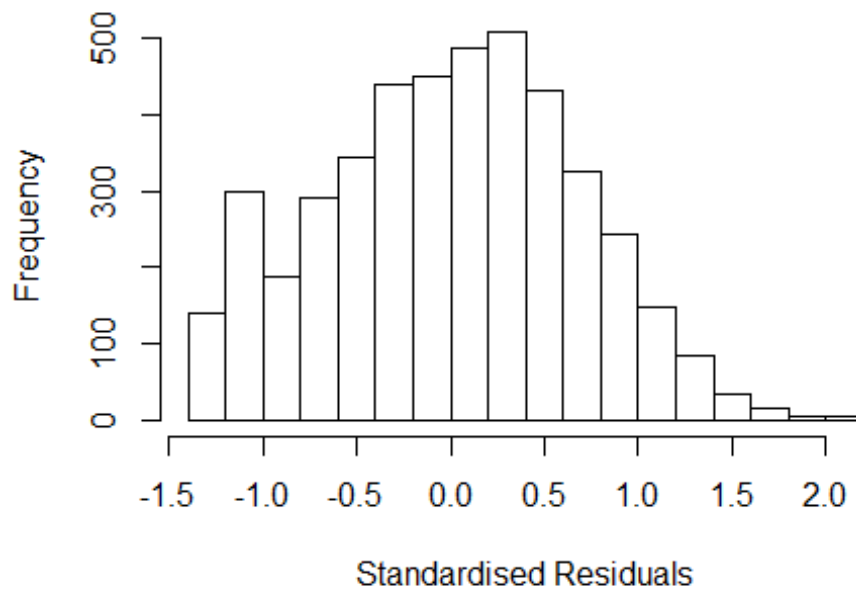
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2886 -0.4841 0.0307 0.4777 2.1969
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20314 0.07509 16.02 <2e-16 ***
## FirstAuthorFemale1 -0.02073 0.02421 -0.86 0.39
## Year1997 0.06499 0.10247 0.63 0.53
## Year1998 0.02814 0.09323 0.30 0.76
## Year1999 0.00355 0.09259 0.04 0.97
## Year2000 -0.01579 0.08851 -0.18 0.86
## Year2001 0.06905 0.09481 0.73 0.47
## Year2002 0.08550 0.09273 0.92 0.36
## Year2003 -0.08479 0.08651 -0.98 0.33
## Year2004 -0.07154 0.08949 -0.80 0.42
## Year2005 -0.04094 0.08576 -0.48 0.63
## Year2006 -0.05459 0.08488 -0.64 0.52
```

```

## Year2007          -0.07000    0.08397   -0.83    0.40
## Year2008          -0.05310    0.08370   -0.63    0.53
## Year2009          -0.05667    0.08302   -0.68    0.49
## Year2010          -0.09391    0.08203   -1.14    0.25
## Year2011           0.02078    0.08387    0.25    0.80
## Year2012          -0.05786    0.08338   -0.69    0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.00591,    Adjusted R-squared:  0.00208
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 355 weights are ~= 1. The remaining 4078 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.878  0.951  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1          1.009
## Year              1.017 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2868 -0.4830 0.0307 0.4787 2.1976
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20185 0.07535 15.95 <2e-16 ***
## LastAuthorFemale1 -0.01295 0.02500 -0.52 0.60
## Year1997 0.06558 0.10258 0.64 0.52
## Year1998 0.02792 0.09332 0.30 0.76
## Year1999 0.00409 0.09266 0.04 0.96
## Year2000 -0.01666 0.08860 -0.19 0.85
## Year2001 0.06888 0.09490 0.73 0.47
## Year2002 0.08496 0.09274 0.92 0.36
## Year2003 -0.08576 0.08655 -0.99 0.32
## Year2004 -0.07203 0.08951 -0.80 0.42
## Year2005 -0.04106 0.08578 -0.48 0.63
## Year2006 -0.05657 0.08489 -0.67 0.51
```

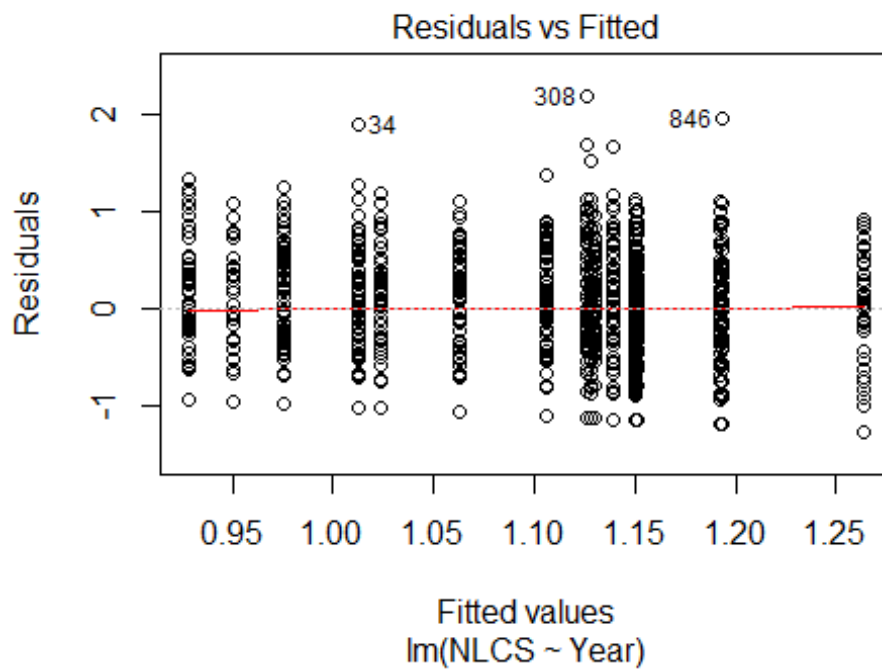
```

## Year2007          -0.07083      0.08395   -0.84      0.40
## Year2008          -0.05385      0.08370   -0.64      0.52
## Year2009          -0.05799      0.08292   -0.70      0.48
## Year2010          -0.09469      0.08205   -1.15      0.25
## Year2011           0.02020      0.08385     0.24      0.81
## Year2012          -0.05935      0.08336   -0.71      0.48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.00581,    Adjusted R-squared:  0.00198
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 343 weights are ~= 1. The remaining 4090 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.319  0.878  0.951   0.916  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4433"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1404"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 121 112 134 114 156 106 94 97 90 114 135 138 116 130 137
## 2011 2012
## 109 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 60 67 34 49 39 47 65 60 66 83 82 61 68 78
## 2011 2012

```



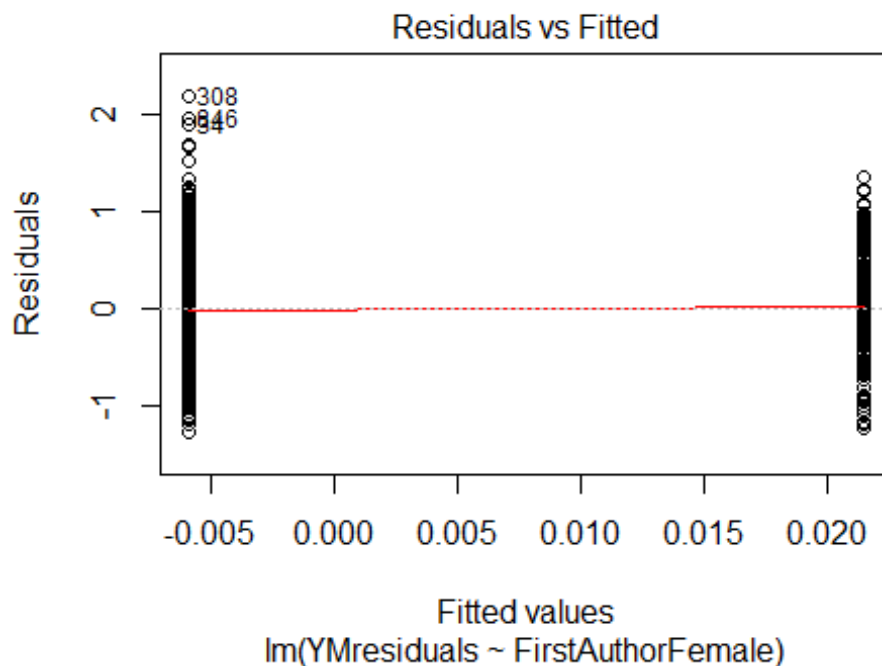
```
## 66 72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 73 56 64 30 49 33 42 59 58 60 76 70 55 64 67
## 2011 2012
## 59 66
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.45, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 2.5438985621441"
## [1] "Male first author team size 2018 geometric mean: 2.00249749194701"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 460, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.44731721855407"
## [1] "Male last author team size 2018 geometric mean: 2.07189836832542"

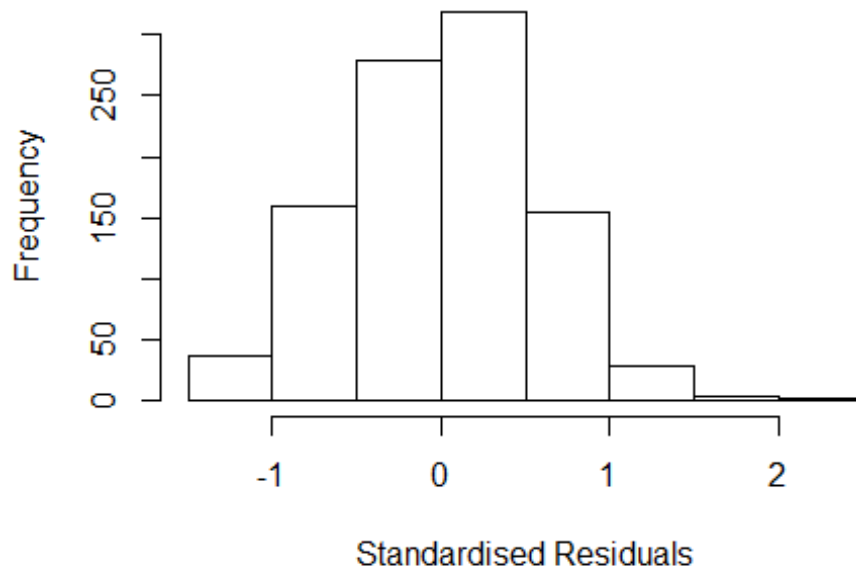
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 360, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.474	1	1.214
LastAuthorFemale	1.496	1	1.223
UniqueAuthors	1.484	4	1.051
Year	1.647	16	1.016

## Residuals from first and last author and team size



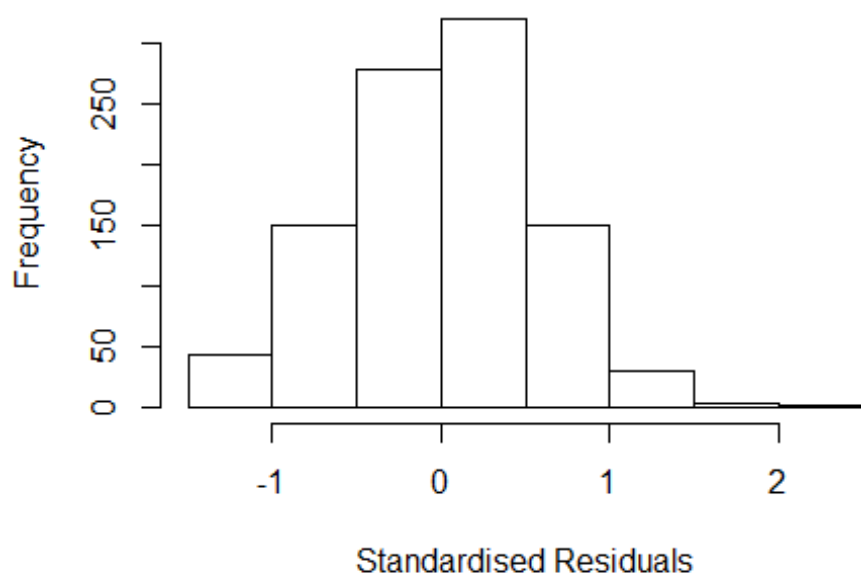
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3144 -0.4140 0.0253 0.3681 2.0923
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9369 0.0658 14.23 < 2e-16 ***
## FirstAuthorFemale1 0.0782 0.0518 1.51 0.13139
## LastAuthorFemale1 -0.0124 0.0514 -0.24 0.80884
## UniqueAuthors2 0.1583 0.0442 3.58 0.00036 ***
## UniqueAuthors3 0.1905 0.0595 3.20 0.00141 **
## UniqueAuthors4 0.2955 0.1131 2.61 0.00916 **
## UniqueAuthors5 0.0992 0.1031 0.96 0.33651
## Year1997 -0.1346 0.1145 -1.18 0.24004
## Year1998 0.0252 0.1097 0.23 0.81846
## Year1999 -0.1403 0.1415 -0.99 0.32164
```

```

## Year2000          0.0919      0.1135      0.81  0.41824
## Year2001          0.2191      0.1270      1.73  0.08484 .
## Year2002          0.1188      0.1289      0.92  0.35695
## Year2003          0.1712      0.1007      1.70  0.08944 .
## Year2004          0.0800      0.1160      0.69  0.49050
## Year2005          0.1065      0.0947      1.13  0.26080
## Year2006          0.0413      0.0956      0.43  0.66577
## Year2007         -0.1427      0.1024     -1.39  0.16361
## Year2008         -0.0854      0.0979     -0.87  0.38361
## Year2009          0.0333      0.0926      0.36  0.71903
## Year2010          0.0442      0.0944      0.47  0.63987
## Year2011          0.0981      0.0936      1.05  0.29496
## Year2012          0.0150      0.0981      0.15  0.87871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.583
## Multiple R-squared:  0.05,   Adjusted R-squared:  0.0282
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 897 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.867  0.951  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.02e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.459 1          1.208
## LastAuthorFemale  1.505 1          1.227
## Year              1.147 16          1.004

```

## Residuals from first and last author



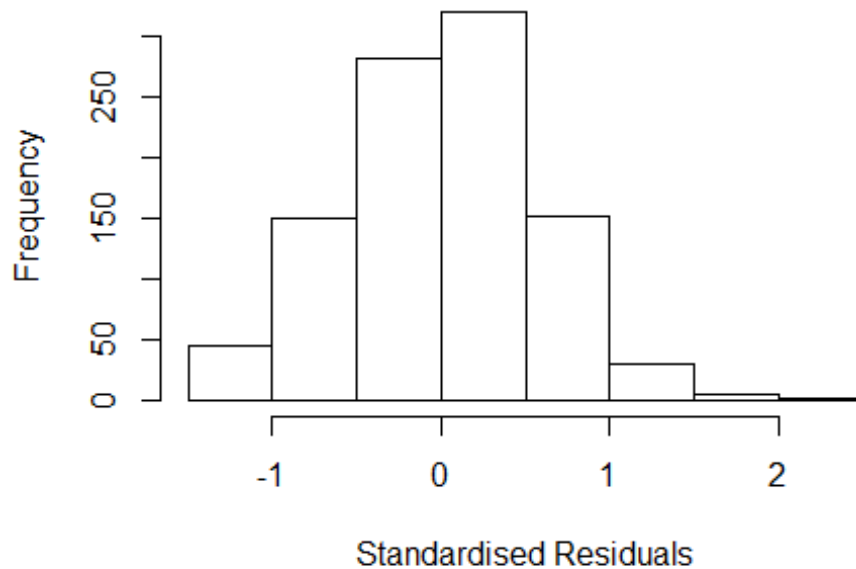
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2364 -0.4047 0.0195 0.3822 2.0109
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99366 0.06615 15.02 <2e-16 ***
## FirstAuthorFemale1 0.07096 0.05286 1.34 0.180
## LastAuthorFemale1 -0.01388 0.05271 -0.26 0.792
## Year1997 -0.14713 0.11758 -1.25 0.211
## Year1998 0.08625 0.10992 0.78 0.433
## Year1999 -0.10572 0.14953 -0.71 0.480
## Year2000 0.12770 0.11623 1.10 0.272
## Year2001 0.24277 0.13026 1.86 0.063 .
## Year2002 0.14343 0.12630 1.14 0.256
## Year2003 0.22588 0.10224 2.21 0.027 *
## Year2004 0.12776 0.11992 1.07 0.287
## Year2005 0.14258 0.09594 1.49 0.138
```

```

## Year2006          0.08715      0.09531      0.91      0.361
## Year2007          -0.06160      0.10174     -0.61      0.545
## Year2008          -0.00701      0.09553     -0.07      0.941
## Year2009           0.09214      0.09038      1.02      0.308
## Year2010           0.09228      0.09235      1.00      0.318
## Year2011           0.13944      0.09298      1.50      0.134
## Year2012           0.09546      0.09458      1.01      0.313
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0282, Adjusted R-squared:  0.01
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 897 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.215  0.869  0.950  0.911  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.07 1          1.034
## Year              1.07 16          1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2352 -0.4090 0.0189 0.3817 2.0124
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99271 0.06608 15.02 <2e-16 ***
## FirstAuthorFemale1 0.06343 0.04530 1.40 0.162
## Year1997 -0.14754 0.11750 -1.26 0.210
## Year1998 0.08492 0.10957 0.78 0.439
## Year1999 -0.10459 0.14983 -0.70 0.485
## Year2000 0.12761 0.11603 1.10 0.272
## Year2001 0.24252 0.13025 1.86 0.063 .
## Year2002 0.14291 0.12606 1.13 0.257
## Year2003 0.22635 0.10220 2.21 0.027 *
## Year2004 0.12654 0.11942 1.06 0.290
## Year2005 0.14146 0.09598 1.47 0.141
## Year2006 0.08530 0.09437 0.90 0.366
```

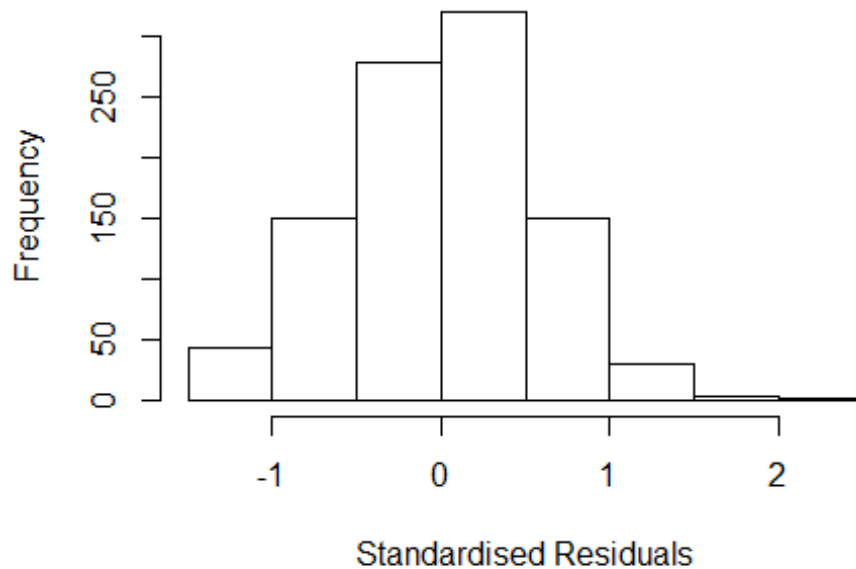
```

## Year2007          -0.06179      0.10173      -0.61      0.544
## Year2008          -0.00765      0.09554      -0.08      0.936
## Year2009           0.09300      0.09037       1.03      0.304
## Year2010           0.09149      0.09191       1.00      0.320
## Year2011           0.13839      0.09275       1.49      0.136
## Year2012           0.09444      0.09409       1.00      0.316
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.011
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 896 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.213  0.868  0.949  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.098 1      1.048
## Year              1.098 16      1.003

```



## Residuals from last author



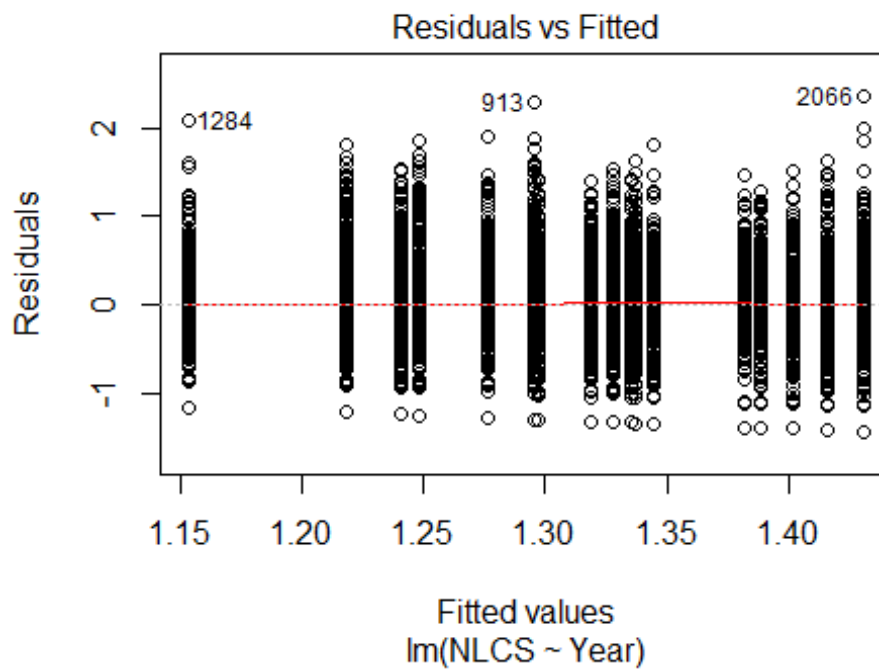
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2405 -0.3985 0.0183 0.3825 1.9980
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9972 0.0662 15.05 <2e-16 ***
## LastAuthorFemale1 0.0247 0.0450 0.55 0.584
## Year1997 -0.1400 0.1179 -1.19 0.235
## Year1998 0.0859 0.1103 0.78 0.436
## Year1999 -0.0999 0.1506 -0.66 0.507
## Year2000 0.1315 0.1158 1.14 0.256
## Year2001 0.2432 0.1312 1.85 0.064 .
## Year2002 0.1527 0.1256 1.22 0.224
## Year2003 0.2293 0.1025 2.24 0.026 *
## Year2004 0.1290 0.1199 1.08 0.282
## Year2005 0.1419 0.0959 1.48 0.139
## Year2006 0.0894 0.0958 0.93 0.351
```

```

## Year2007          -0.0556      0.1019    -0.55      0.586
## Year2008          -0.0011      0.0956    -0.01      0.991
## Year2009           0.0984      0.0903      1.09      0.276
## Year2010           0.0962      0.0924      1.04      0.298
## Year2011           0.1399      0.0930      1.50      0.133
## Year2012           0.1035      0.0947      1.09      0.275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0264, Adjusted R-squared:  0.00921
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 889 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.221  0.864  0.948  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 981"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1405"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  317  342  362  337  378  374  314  249  238  311  317  381  350  342  371
## 2011 2012
##  380  340
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  251  271  289  268  289  256  251  205  190  262  259  312  291  292  309
## 2011 2012

```

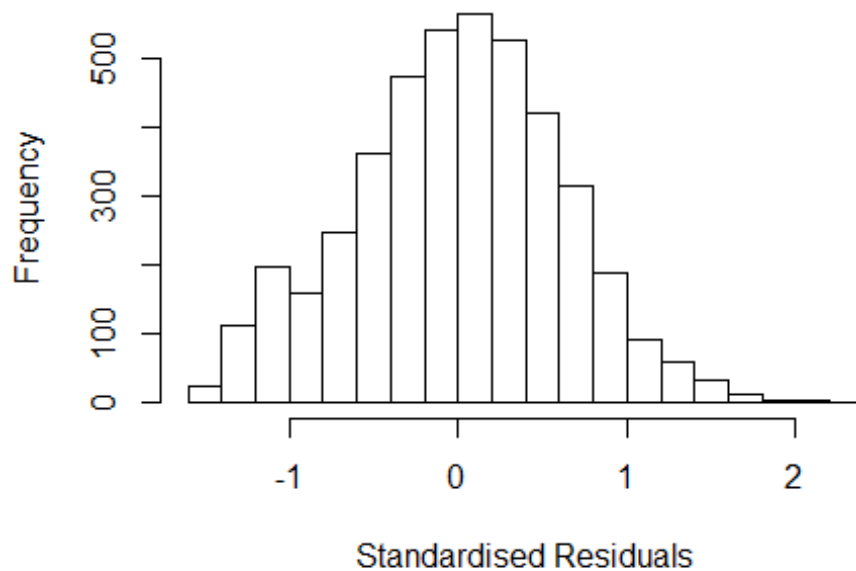
```
## 315 286
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 241 260 275 263 275 243 243 196 175 241 242 286 279 268 279
## 2011 2012
## 299 267
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 19, df = 1, p-value = 1e-05
```



## Residuals from first and last author and team size



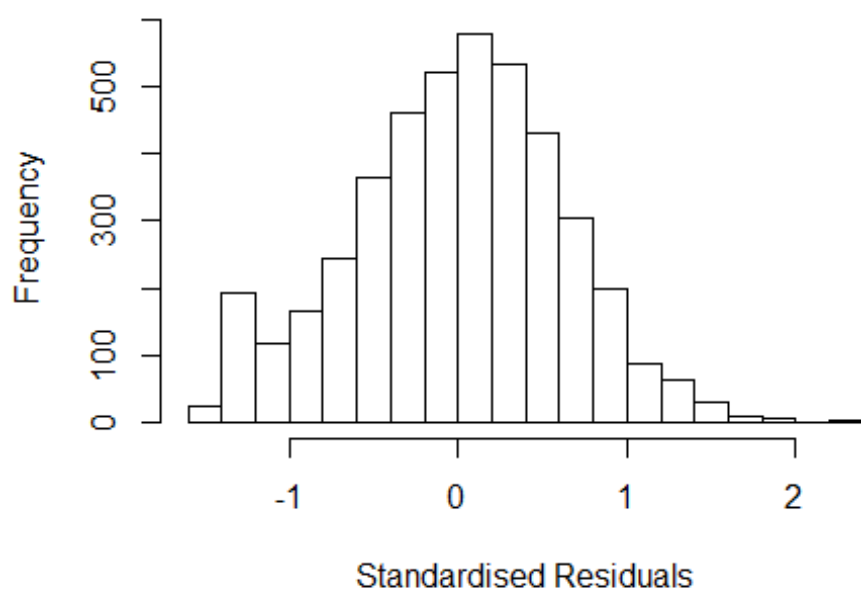
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.600 -0.404 0.016 0.416 2.257
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14779 0.05700 20.14 < 2e-16 ***
## FirstAuthorFemale1 -0.00153 0.02433 -0.06 0.9498
## LastAuthorFemale1 0.00879 0.02467 0.36 0.7217
## UniqueAuthors2 0.18019 0.02236 8.06 1.0e-15 ***
## UniqueAuthors3 0.25260 0.02800 9.02 < 2e-16 ***
## UniqueAuthors4 0.25591 0.04613 5.55 3.1e-08 ***
## UniqueAuthors5 0.18355 0.09072 2.02 0.0431 *
## Year1997 -0.02370 0.07546 -0.31 0.7535
## Year1998 0.04511 0.07332 0.62 0.5385
## Year1999 -0.06180 0.07041 -0.88 0.3802
```

```

## Year2000      0.09609      0.06670      1.44      0.1498
## Year2001      0.19924      0.07274      2.74      0.0062 **
## Year2002      0.09784      0.07071      1.38      0.1666
## Year2003      0.14101      0.06806      2.07      0.0383 *
## Year2004      0.13916      0.07050      1.97      0.0485 *
## Year2005      0.16605      0.06729      2.47      0.0136 *
## Year2006      0.12376      0.06474      1.91      0.0560 .
## Year2007      0.08738      0.06504      1.34      0.1792
## Year2008      0.04640      0.06498      0.71      0.4752
## Year2009     -0.03560      0.06913     -0.52      0.6066
## Year2010      0.04156      0.06789      0.61      0.5404
## Year2011      0.05468      0.06974      0.78      0.4331
## Year2012      0.01600      0.06873      0.23      0.8160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.612
## Multiple R-squared:  0.0412, Adjusted R-squared:  0.0363
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 377 weights are ~= 1. The remaining 3955 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.144  0.864  0.952  0.904  0.984  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.31e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.335 1           1.156
## LastAuthorFemale  1.337 1           1.156
## Year              1.027 16           1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4478 -0.4109 0.0269 0.4174 2.3550
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22616 0.05776 21.23 <2e-16 ***
## FirstAuthorFemale1 0.00765 0.02454 0.31 0.7554
## LastAuthorFemale1 0.01106 0.02496 0.44 0.6576
## Year1997 -0.02642 0.07775 -0.34 0.7341
## Year1998 0.05853 0.07476 0.78 0.4337
## Year1999 -0.06912 0.07192 -0.96 0.3366
## Year2000 0.11739 0.06808 1.72 0.0847 .
## Year2001 0.20288 0.07362 2.76 0.0059 **
## Year2002 0.12600 0.07217 1.75 0.0809 .
## Year2003 0.16066 0.06908 2.33 0.0201 *
## Year2004 0.17634 0.07216 2.44 0.0146 *
## Year2005 0.18572 0.06864 2.71 0.0068 **
```

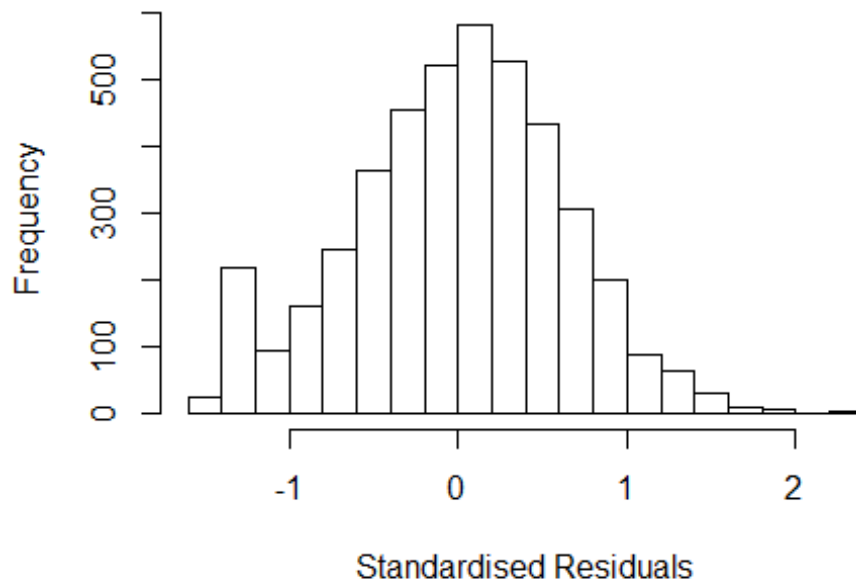
```

## Year2006          0.15408    0.06594    2.34    0.0195 *
## Year2007          0.11956    0.06593    1.81    0.0699 .
## Year2008          0.07188    0.06605    1.09    0.2765
## Year2009         -0.00800    0.07066   -0.11    0.9098
## Year2010          0.09667    0.06866    1.41    0.1592
## Year2011          0.10420    0.07141    1.46    0.1446
## Year2012          0.05913    0.06993    0.85    0.3978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0142, Adjusted R-squared:  0.0101
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 341 weights are ~= 1. The remaining 3991 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.113  0.870   0.951   0.904   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.31e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.015 1      1.008
## Year              1.015 16      1.000

```



## Residuals from first author



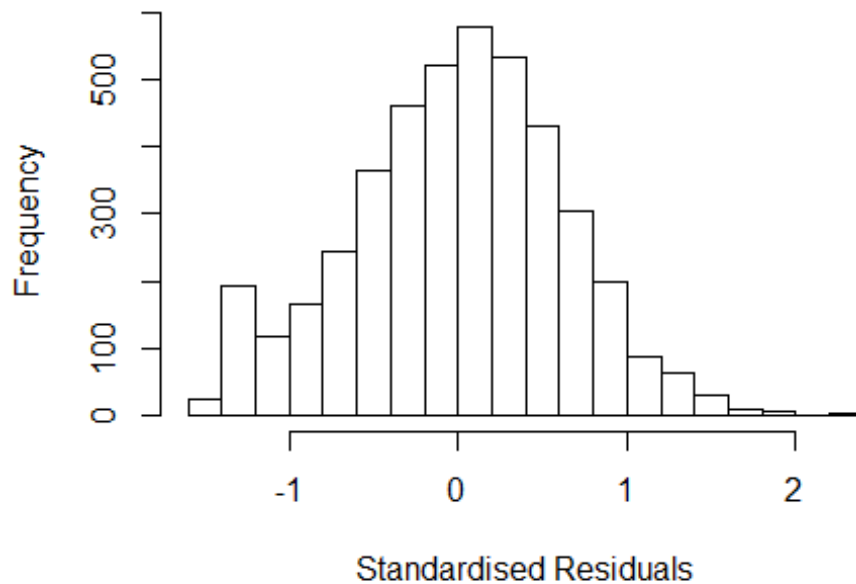
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4434 -0.4105 0.0271 0.4191 2.3538
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22724 0.05777 21.24 <2e-16 ***
## FirstAuthorFemale1 0.01315 0.02140 0.61 0.5388
## Year1997 -0.02639 0.07781 -0.34 0.7345
## Year1998 0.05891 0.07478 0.79 0.4308
## Year1999 -0.06892 0.07194 -0.96 0.3381
## Year2000 0.11799 0.06806 1.73 0.0831 .
## Year2001 0.20296 0.07367 2.75 0.0059 **
## Year2002 0.12629 0.07220 1.75 0.0803 .
## Year2003 0.16080 0.06911 2.33 0.0200 *
## Year2004 0.17686 0.07213 2.45 0.0142 *
## Year2005 0.18607 0.06867 2.71 0.0068 **
## Year2006 0.15453 0.06595 2.34 0.0192 *
```

```

## Year2007          0.11983      0.06597      1.82      0.0694 .
## Year2008          0.07195      0.06609      1.09      0.2763
## Year2009         -0.00748      0.07069     -0.11      0.9157
## Year2010          0.09651      0.06869      1.40      0.1601
## Year2011          0.10477      0.07145      1.47      0.1426
## Year2012          0.05966      0.06996      0.85      0.3939
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0142, Adjusted R-squared:  0.0103
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 335 weights are ~= 1. The remaining 3997 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.113  0.870   0.951   0.904   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.31e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1      1.008
## Year              1.017 16      1.001

```

## Residuals from last author



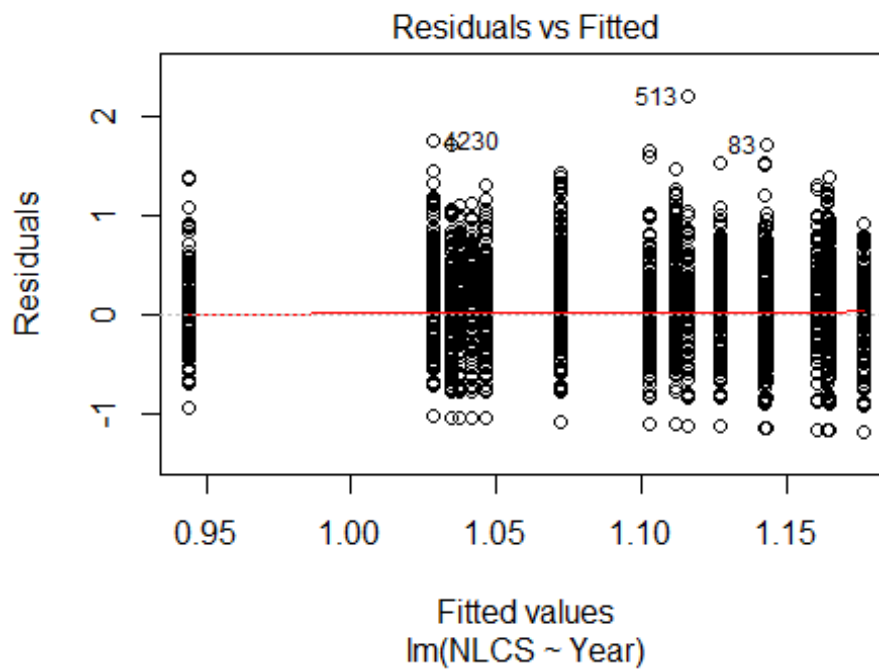
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4450 -0.4101 0.0259 0.4172 2.3539
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2267 0.0577 21.25 <2e-16 ***
## LastAuthorFemale1 0.0149 0.0217 0.69 0.4920
## Year1997 -0.0262 0.0777 -0.34 0.7362
## Year1998 0.0589 0.0747 0.79 0.4303
## Year1999 -0.0687 0.0719 -0.96 0.3393
## Year2000 0.1177 0.0681 1.73 0.0838 .
## Year2001 0.2034 0.0736 2.76 0.0057 **
## Year2002 0.1263 0.0722 1.75 0.0802 .
## Year2003 0.1610 0.0691 2.33 0.0198 *
## Year2004 0.1766 0.0722 2.45 0.0144 *
## Year2005 0.1863 0.0686 2.72 0.0066 **
## Year2006 0.1545 0.0659 2.34 0.0191 *
```

```

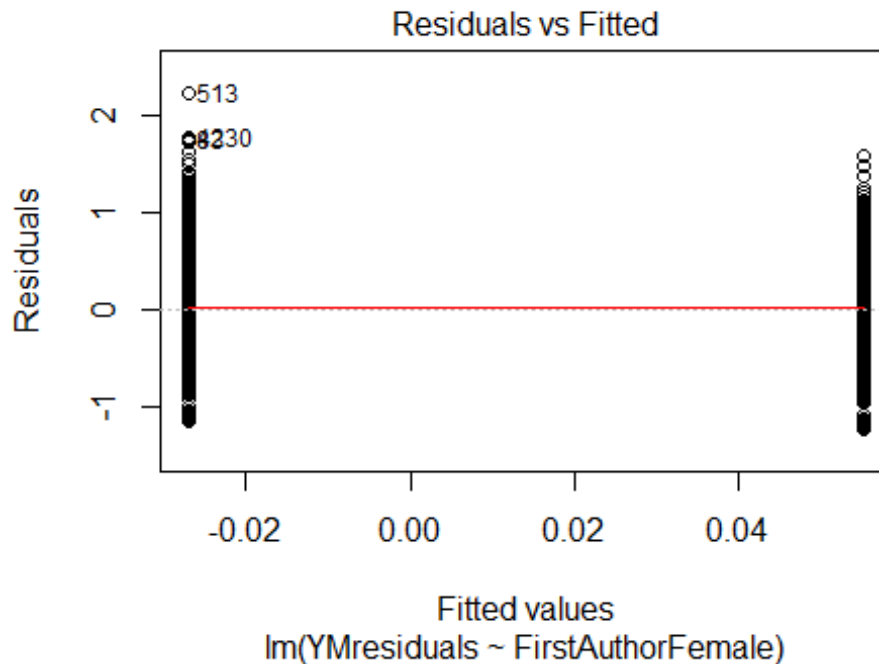
## Year2007          0.1200      0.0659      1.82      0.0686 .
## Year2008          0.0726      0.0660      1.10      0.2716
## Year2009         -0.0070      0.0705     -0.10      0.9210
## Year2010          0.0974      0.0687      1.42      0.1558
## Year2011          0.1048      0.0713      1.47      0.1419
## Year2012          0.0595      0.0699      0.85      0.3945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0142, Adjusted R-squared:  0.0103
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 345 weights are ~= 1. The remaining 3987 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.114  0.870  0.951  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.31e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4332"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1406"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 169 160 162 169 208 176 202 188 193 218 236 273 253 333 333
## 2011 2012
## 310 308
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 127 111 101 113 108 118 144 144 153 164 179 207 194 245 264
## 2011 2012

```

```
## 244 254
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 123 104 96 107 102 106 135 134 147 154 164 191 175 228 250
## 2011 2012
## 230 236
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 48, df = 16, p-value = 5e-05
```

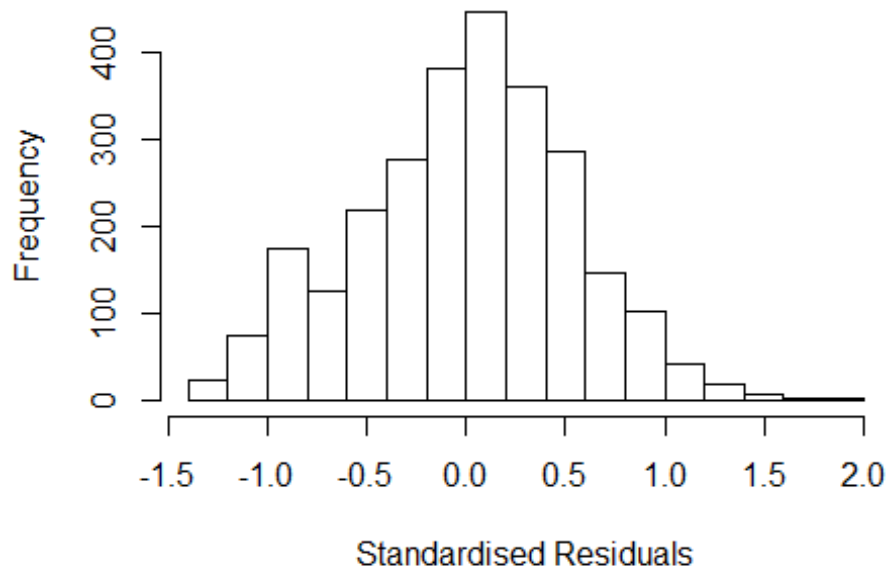


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.7, df = 1, p-value = 0.009
```



```
## [1] "Female first author team size 2018 geometric mean: 2.02406085259791"
## [1] "Male first author team size 2018 geometric mean: 2.05779497228691"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3800, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.1430295469548"
## [1] "Male last author team size 2018 geometric mean: 1.98735397380225"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.276 1      1.130
## LastAuthorFemale  1.236 1      1.112
## UniqueAuthors     1.169 4      1.020
## Year              1.200 16      1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3482 -0.3662  0.0298  0.3567  1.9090
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.017304   0.050209   20.26 < 2e-16 ***
## FirstAuthorFemale1 0.054488   0.024721    2.20 0.02760 *
## LastAuthorFemale1 0.017515   0.024788    0.71 0.47988
## UniqueAuthors2    0.217023   0.024795    8.75 < 2e-16 ***
## UniqueAuthors3    0.268102   0.032994    8.13 6.7e-16 ***
## UniqueAuthors4    0.291296   0.051027    5.71 1.3e-08 ***
## UniqueAuthors5    0.167820   0.076333    2.20 0.02800 *
## Year1997          -0.001505   0.074596   -0.02 0.98390
## Year1998          -0.052261   0.076755   -0.68 0.49600
## Year1999          -0.000407   0.072174   -0.01 0.99551
```

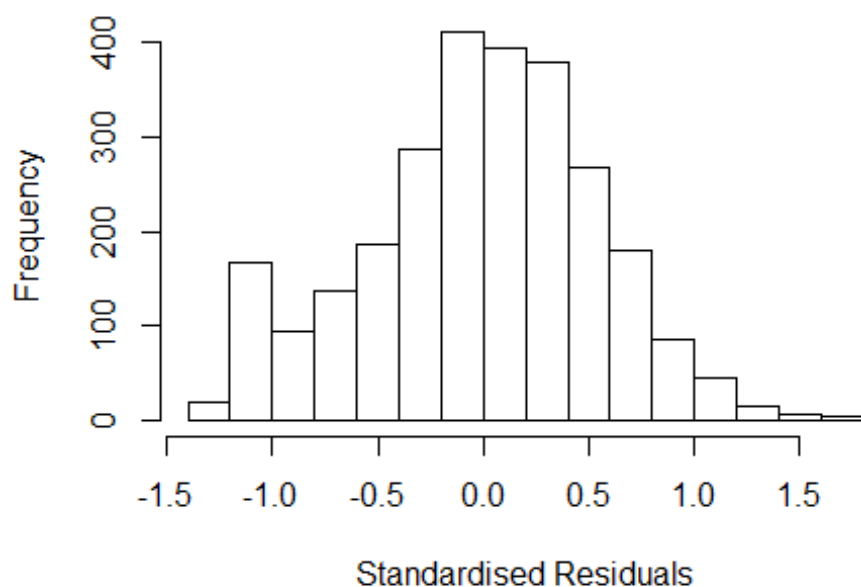
```

## Year2000      -0.018508    0.081269   -0.23  0.81986
## Year2001      0.041852    0.071293    0.59  0.55722
## Year2002     -0.113322    0.070688   -1.60  0.10902
## Year2003     -0.134760    0.066373   -2.03  0.04242 *
## Year2004     -0.235123    0.062032   -3.79  0.00015 ***
## Year2005     -0.149367    0.065694   -2.27  0.02306 *
## Year2006     -0.148234    0.061595   -2.41  0.01617 *
## Year2007     -0.089450    0.059078   -1.51  0.13012
## Year2008     -0.060295    0.062124   -0.97  0.33186
## Year2009      0.011032    0.065570    0.17  0.86640
## Year2010     -0.095716    0.064335   -1.49  0.13692
## Year2011     -0.153332    0.064443   -2.38  0.01741 *
## Year2012     -0.088673    0.064708   -1.37  0.17069
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.0645, Adjusted R-squared:  0.0567
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 231 weights are ~= 1. The remaining 2451 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.145  0.855  0.947  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.252 1      1.119
## LastAuthorFemale  1.230 1      1.109
## Year              1.061 16      1.002

```



## Residuals from first and last author



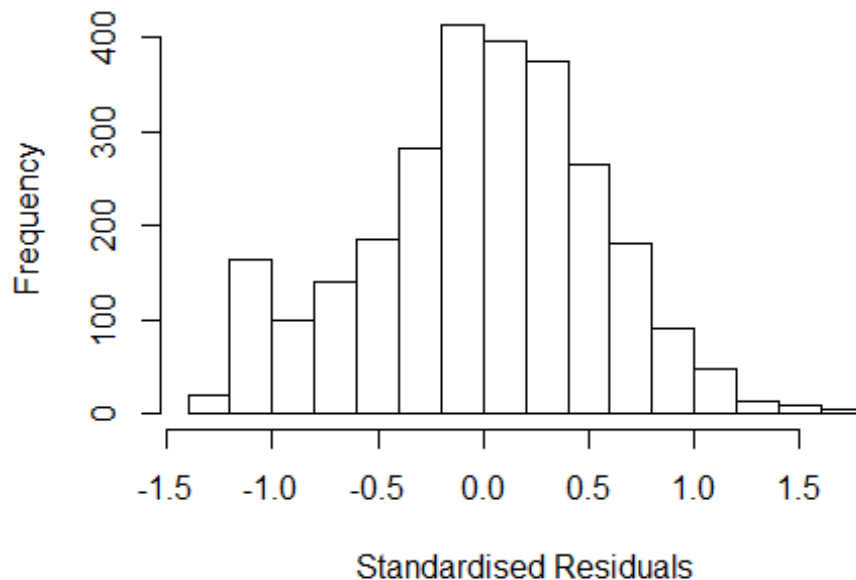
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2729 -0.3526 0.0221 0.3626 1.7778
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1112 0.0489 22.72 < 2e-16 ***
## FirstAuthorFemale1 0.0767 0.0250 3.07 0.00220 **
## LastAuthorFemale1 0.0247 0.0253 0.98 0.32901
## Year1997 0.0461 0.0738 0.62 0.53230
## Year1998 -0.0450 0.0743 -0.61 0.54460
## Year1999 0.0234 0.0713 0.33 0.74245
## Year2000 0.0200 0.0785 0.25 0.79921
## Year2001 0.0603 0.0706 0.85 0.39305
## Year2002 -0.0951 0.0721 -1.32 0.18732
## Year2003 -0.1100 0.0665 -1.65 0.09813 .
## Year2004 -0.2158 0.0631 -3.42 0.00063 ***
## Year2005 -0.1259 0.0657 -1.92 0.05553 .
```

```

## Year2006          -0.1086      0.0628    -1.73  0.08388 .
## Year2007          -0.0503      0.0585    -0.86  0.38939
## Year2008          -0.0163      0.0614    -0.27  0.79085
## Year2009           0.0303      0.0660     0.46  0.64640
## Year2010          -0.0692      0.0641    -1.08  0.28006
## Year2011          -0.1161      0.0655    -1.77  0.07624 .
## Year2012          -0.0334      0.0647    -0.52  0.60586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.0206, Adjusted R-squared:  0.0139
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 233 weights are ~= 1. The remaining 2449 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.234  0.855  0.950  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1          1.022
## Year              1.044 16          1.001

```

## Residuals from first author



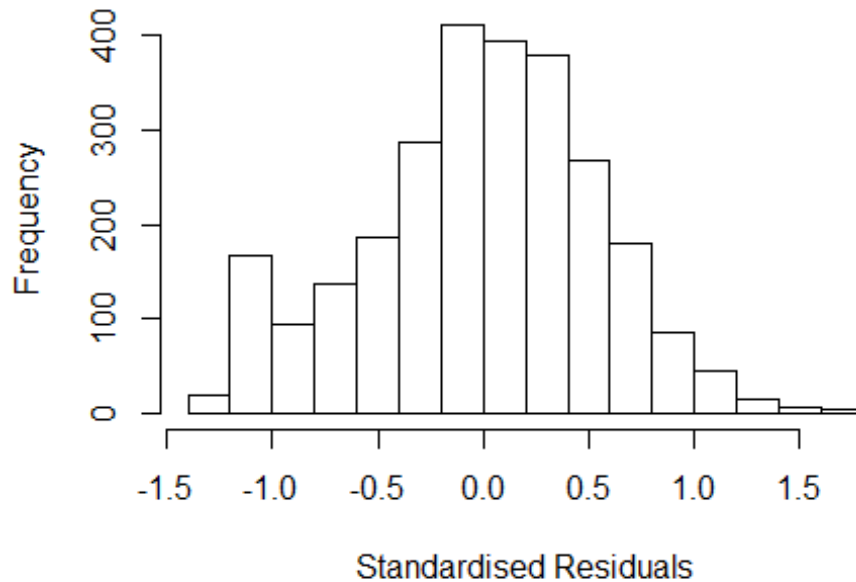
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.264 -0.350 0.023 0.362 1.774
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1147 0.0487 22.91 < 2e-16 ***
## FirstAuthorFemale1 0.0884 0.0229 3.86 0.00012 ***
## Year1997 0.0474 0.0739 0.64 0.52100
## Year1998 -0.0447 0.0742 -0.60 0.54653
## Year1999 0.0227 0.0712 0.32 0.74973
## Year2000 0.0209 0.0785 0.27 0.78978
## Year2001 0.0614 0.0703 0.87 0.38291
## Year2002 -0.0941 0.0721 -1.30 0.19201
## Year2003 -0.1088 0.0663 -1.64 0.10094
## Year2004 -0.2141 0.0631 -3.40 0.00069 ***
## Year2005 -0.1257 0.0656 -1.92 0.05556 .
## Year2006 -0.1094 0.0627 -1.74 0.08117 .
```

```

## Year2007          -0.0498      0.0584   -0.85   0.39369
## Year2008          -0.0158      0.0613   -0.26   0.79604
## Year2009           0.0309      0.0658    0.47   0.63825
## Year2010          -0.0679      0.0640   -1.06   0.28829
## Year2011          -0.1157      0.0653   -1.77   0.07656 .
## Year2012          -0.0324      0.0645   -0.50   0.61579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.0202, Adjusted R-squared:  0.0139
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 236 weights are ~= 1. The remaining 2446 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.237  0.857  0.949  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.027 1      1.013
## Year              1.027 16      1.001

```

## Residuals from last author



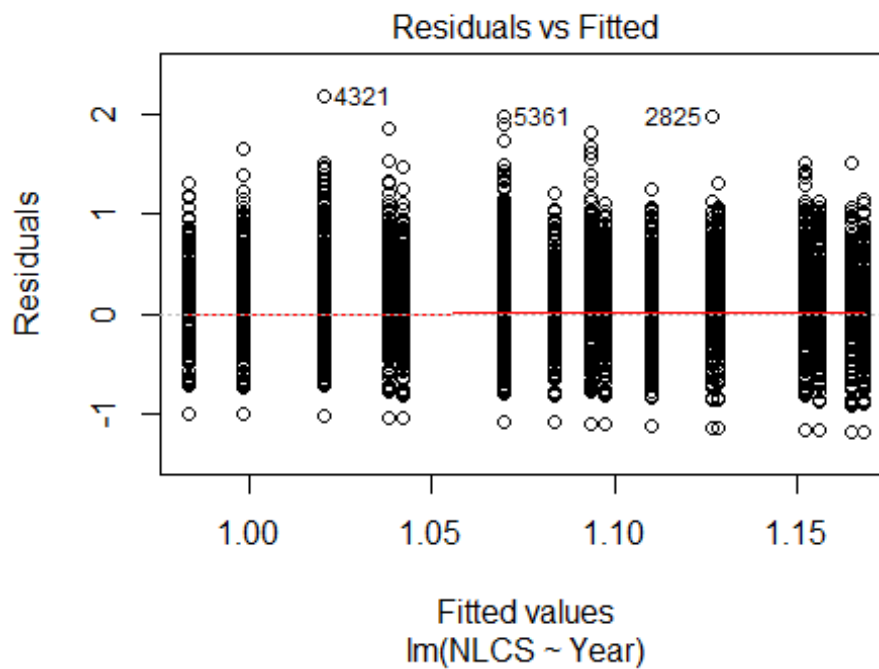
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.243 -0.346 0.028 0.361 1.762
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1225 0.0489 22.94 < 2e-16 ***
## LastAuthorFemale1 0.0612 0.0231 2.65 0.00802 **
## Year1997 0.0392 0.0732 0.54 0.59207
## Year1998 -0.0490 0.0743 -0.66 0.50945
## Year1999 0.0248 0.0717 0.35 0.72981
## Year2000 0.0159 0.0782 0.20 0.83892
## Year2001 0.0597 0.0711 0.84 0.40130
## Year2002 -0.0917 0.0726 -1.26 0.20646
## Year2003 -0.1045 0.0665 -1.57 0.11631
## Year2004 -0.2140 0.0632 -3.38 0.00072 ***
## Year2005 -0.1258 0.0656 -1.92 0.05534 .
## Year2006 -0.1014 0.0630 -1.61 0.10780
```

```

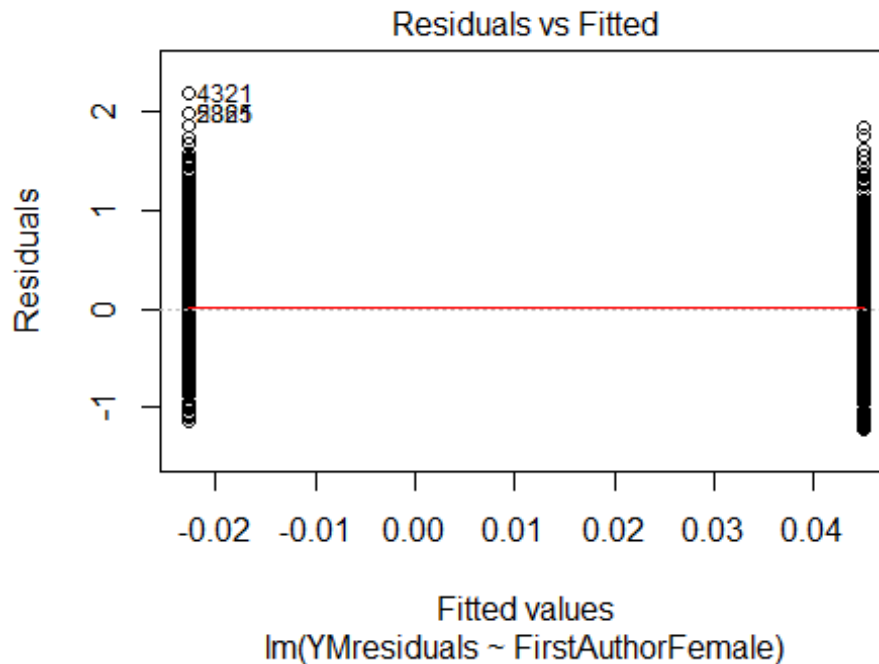
## Year2007          -0.0442      0.0584   -0.76  0.44953
## Year2008          -0.0129      0.0614   -0.21  0.83339
## Year2009           0.0355      0.0663    0.54  0.59220
## Year2010          -0.0645      0.0640   -1.01  0.31361
## Year2011          -0.1113      0.0660   -1.69  0.09178 .
## Year2012          -0.0300      0.0651   -0.46  0.64483
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.0172, Adjusted R-squared:  0.0109
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 239 weights are ~ = 1. The remaining 2443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.242  0.856  0.949   0.898  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2682"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1407"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 239 258 236 232 334 267 257 255 279 201 261 327 289 401 436
## 2011 2012
## 400 399
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 218 229 211 209 292 227 223 228 243 186 232 285 266 347 375
## 2011 2012

```

```
## 352 343
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 211 216 208 205 279 217 213 215 229 177 221 275 257 332 360
## 2011 2012
## 325 331
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 96, df = 16, p-value = 2e-13
```



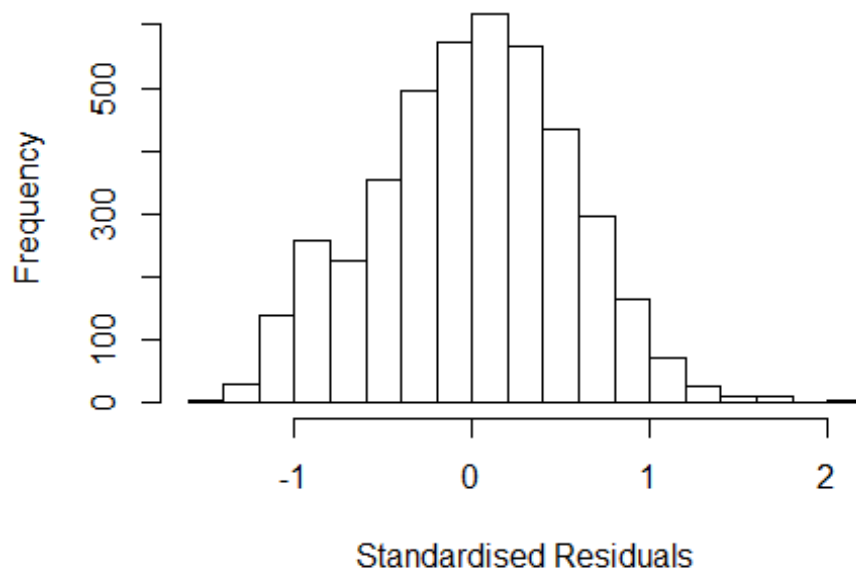
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.3, df = 1, p-value = 0.007
```



```
## [1] "Female first author team size 2018 geometric mean: 1.80485840108111"
## [1] "Male first author team size 2018 geometric mean: 1.79163045528052"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6400, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.77769997798912"
## [1] "Male last author team size 2018 geometric mean: 1.8111792701094"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6200, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.477 1      1.215
## LastAuthorFemale  1.478 1      1.216
## UniqueAuthors    1.121 4      1.014
## Year             1.153 16     1.004
```



## Residuals from first and last author and team size



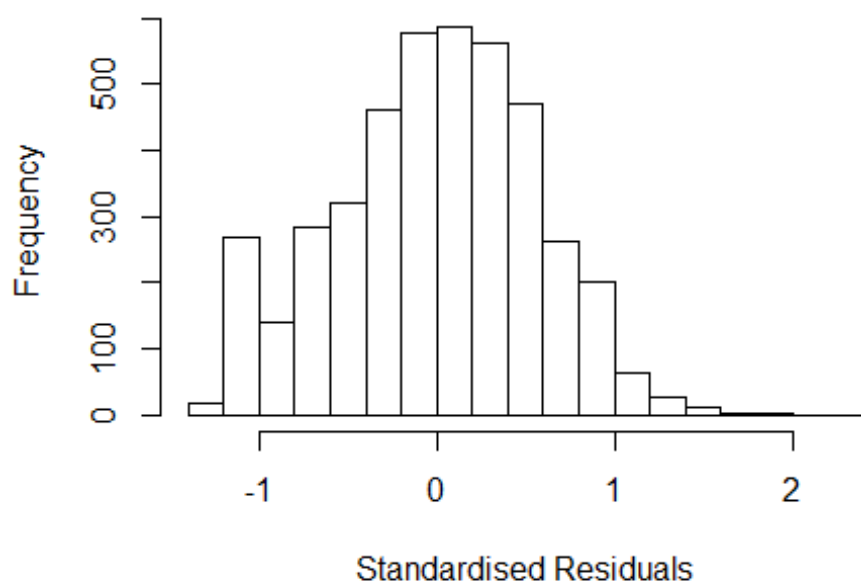
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4559 -0.3672 0.0249 0.3845 2.1204
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.872525 0.034871 25.02 < 2e-16 ***
## FirstAuthorFemale1 0.047965 0.022123 2.17 0.03020 *
## LastAuthorFemale1 0.024147 0.022469 1.07 0.28258
## UniqueAuthors2 0.230894 0.020357 11.34 < 2e-16 ***
## UniqueAuthors3 0.224227 0.026757 8.38 < 2e-16 ***
## UniqueAuthors4 0.331739 0.041767 7.94 2.5e-15 ***
## UniqueAuthors5 0.404933 0.080817 5.01 5.6e-07 ***
## Year1997 0.035730 0.053520 0.67 0.50443
## Year1998 0.000308 0.050765 0.01 0.99517
## Year1999 0.089844 0.047525 1.89 0.05876 .
```

```

## Year2000      0.035943    0.046328    0.78  0.43789
## Year2001      0.137644    0.051730    2.66  0.00782 **
## Year2002      0.152696    0.047211    3.23  0.00123 **
## Year2003      0.146353    0.048872    2.99  0.00276 **
## Year2004      0.064713    0.050799    1.27  0.20277
## Year2005      0.103066    0.054202    1.90  0.05730 .
## Year2006      0.169949    0.049070    3.46  0.00054 ***
## Year2007      0.101943    0.046888    2.17  0.02975 *
## Year2008      0.106874    0.048657    2.20  0.02811 *
## Year2009      0.021346    0.052599    0.41  0.68489
## Year2010      0.068268    0.048110    1.42  0.15597
## Year2011      0.057531    0.051289    1.12  0.26205
## Year2012      0.130450    0.051249    2.55  0.01095 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.059, Adjusted R-squared:  0.0541
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 328 weights are ~= 1. The remaining 3943 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.114  0.870  0.951  0.908  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.478 1      1.216
## LastAuthorFemale  1.484 1      1.218
## Year              1.041 16      1.001

```

## Residuals from first and last author



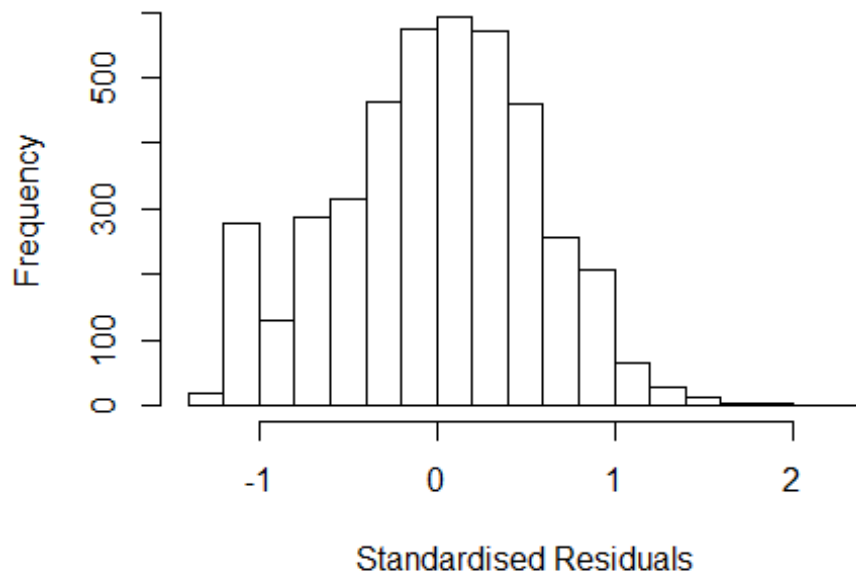
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2368 -0.3819 0.0211 0.3944 2.2073
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9546 0.0353 27.07 < 2e-16 ***
## FirstAuthorFemale1 0.0558 0.0226 2.46 0.01378 *
## LastAuthorFemale1 0.0268 0.0230 1.16 0.24463
## Year1997 0.0433 0.0539 0.80 0.42133
## Year1998 0.0129 0.0524 0.25 0.80517
## Year1999 0.1024 0.0489 2.09 0.03631 *
## Year2000 0.0520 0.0472 1.10 0.27133
## Year2001 0.1655 0.0520 3.18 0.00148 **
## Year2002 0.1837 0.0490 3.75 0.00018 ***
## Year2003 0.1753 0.0496 3.54 0.00041 ***
## Year2004 0.0735 0.0524 1.40 0.16114
## Year2005 0.1368 0.0558 2.45 0.01422 *
```

```

## Year2006          0.1997      0.0498      4.01  6.2e-05 ***
## Year2007          0.1381      0.0480      2.87  0.00407 **
## Year2008          0.1284      0.0490      2.62  0.00885 **
## Year2009          0.0271      0.0539      0.50  0.61521
## Year2010          0.1097      0.0497      2.21  0.02750 *
## Year2011          0.0863      0.0525      1.64  0.10017
## Year2012          0.1702      0.0520      3.27  0.00107 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.015, Adjusted R-squared:  0.0108
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 377 weights are ~= 1. The remaining 3894 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0958 0.8660 0.9500 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year      1.025 16      1.001

```

## Residuals from first author



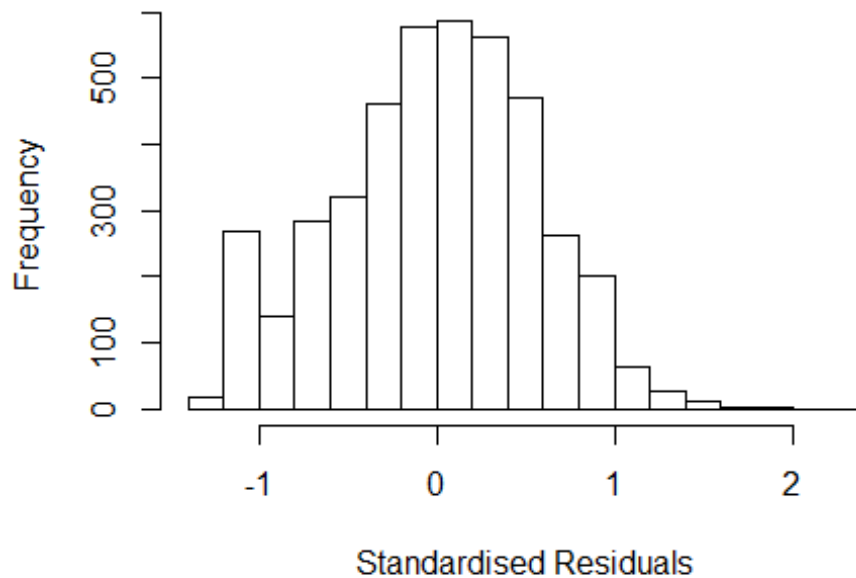
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2287 -0.3795  0.0242  0.3917  2.2039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9576    0.0352   27.21 < 2e-16 ***
## FirstAuthorFemale1 0.0715    0.0188    3.79 0.00015 ***
## Year1997        0.0453    0.0538    0.84 0.40021
## Year1998        0.0139    0.0524    0.27 0.79065
## Year1999        0.1021    0.0489    2.09 0.03682 *
## Year2000        0.0528    0.0473    1.12 0.26387
## Year2001        0.1646    0.0520    3.17 0.00156 **
## Year2002        0.1842    0.0491    3.75 0.00018 ***
## Year2003        0.1753    0.0496    3.54 0.00041 ***
## Year2004        0.0747    0.0524    1.43 0.15407
## Year2005        0.1377    0.0558    2.47 0.01366 *
## Year2006        0.1996    0.0498    4.01 6.3e-05 ***
```

```

## Year2007          0.1381      0.0481      2.87  0.00407 **
## Year2008          0.1283      0.0491      2.61  0.00902 **
## Year2009          0.0275      0.0540      0.51  0.61049
## Year2010          0.1104      0.0497      2.22  0.02656 *
## Year2011          0.0870      0.0525      1.66  0.09758 .
## Year2012          0.1730      0.0520      3.33  0.00088 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.0146, Adjusted R-squared:  0.0107
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 365 weights are ~= 1. The remaining 3906 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.097  0.868  0.950  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.014
## Year            1.029 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2247 -0.3826 0.0229 0.3921 2.1988
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9604 0.0352 27.28 < 2e-16 ***
## LastAuthorFemale1 0.0600 0.0191 3.13 0.00174 **
## Year1997 0.0431 0.0542 0.79 0.42683
## Year1998 0.0108 0.0525 0.21 0.83730
## Year1999 0.1055 0.0489 2.16 0.03120 *
## Year2000 0.0541 0.0472 1.15 0.25199
## Year2001 0.1683 0.0521 3.23 0.00126 **
## Year2002 0.1855 0.0489 3.79 0.00015 ***
## Year2003 0.1772 0.0497 3.56 0.00037 ***
## Year2004 0.0752 0.0525 1.43 0.15229
## Year2005 0.1414 0.0557 2.54 0.01116 *
## Year2006 0.2044 0.0498 4.10 4.1e-05 ***
```

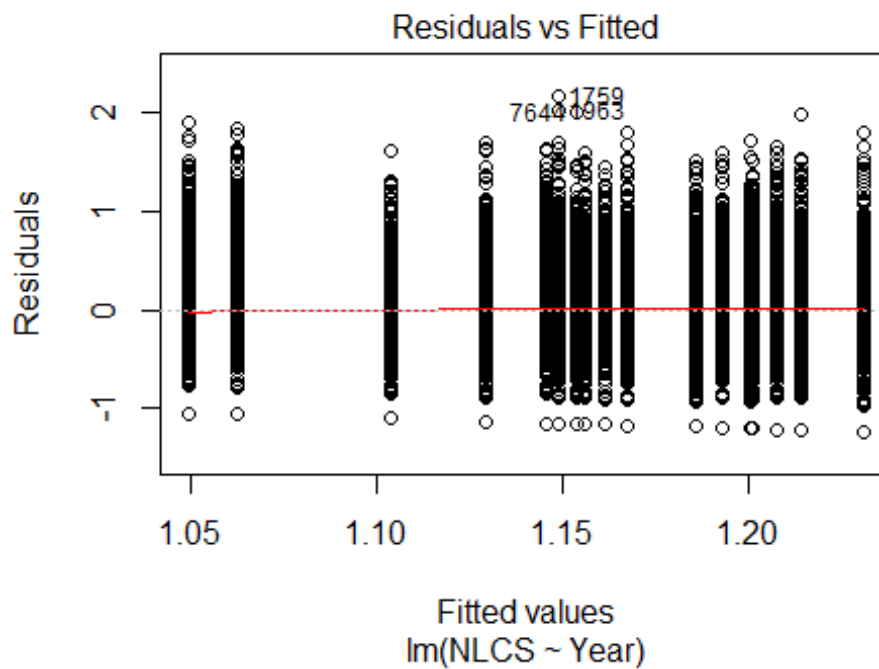
```

## Year2007          0.1404      0.0481      2.92  0.00352 **
## Year2008          0.1325      0.0491      2.70  0.00703 **
## Year2009          0.0298      0.0540      0.55  0.58073
## Year2010          0.1117      0.0499      2.24  0.02532 *
## Year2011          0.0889      0.0525      1.69  0.09047 .
## Year2012          0.1725      0.0519      3.32  0.00090 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.0136, Adjusted R-squared:  0.0097
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 376 weights are ~= 1. The remaining 3895 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0994 0.8670 0.9490 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4271"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1408"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 623 581 591 547 665 570 571 497 497 572 614 753 669 785 712
## 2011 2012
## 735 697
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 470 425 407 379 443 385 423 379 399 451 482 587 524 620 564
## 2011 2012

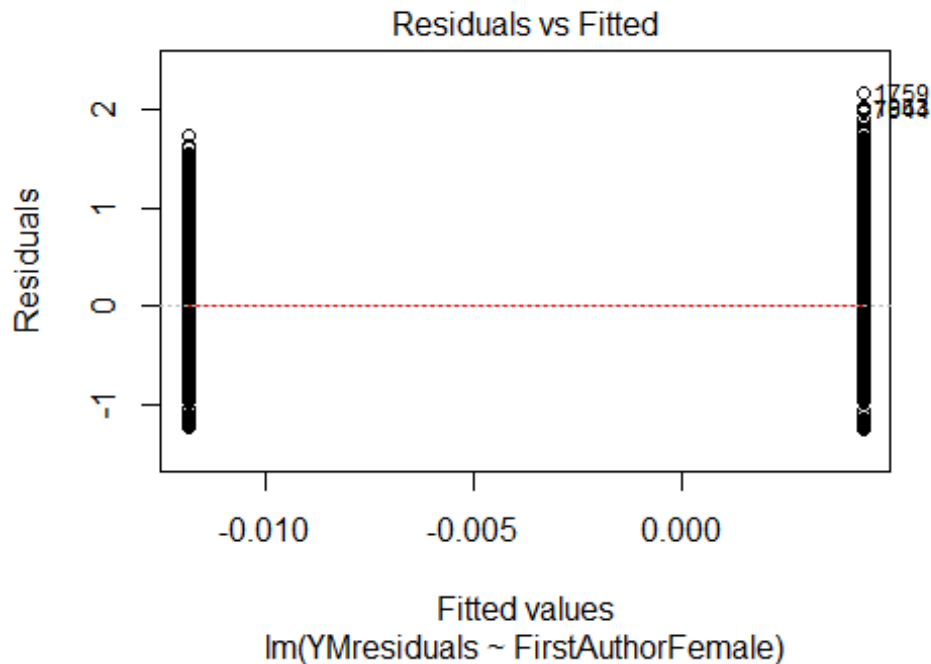
```



```
## 574 569
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 452 404 389 367 420 366 400 360 374 423 460 554 494 581 514
## 2011 2012
## 543 527
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 3e-06
```

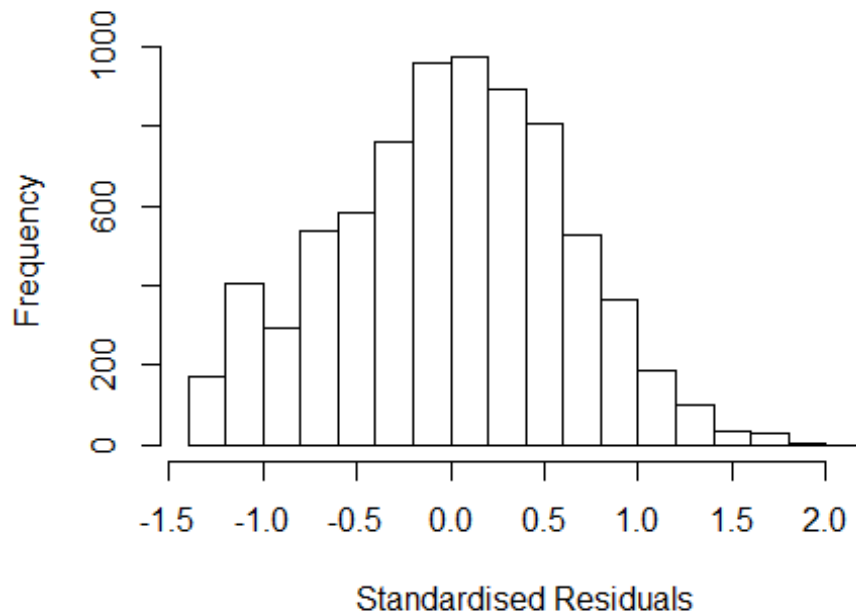


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.22, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 1.88924671382886"
## [1] "Male first author team size 2018 geometric mean: 1.98183659320214"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 36000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.8268770180535"
## [1] "Male last author team size 2018 geometric mean: 2.00721782763511"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.285 1      1.134
## LastAuthorFemale  1.284 1      1.133
## UniqueAuthors     1.061 4      1.007
## Year              1.072 16      1.002
```

## Residuals from first and last author and team size



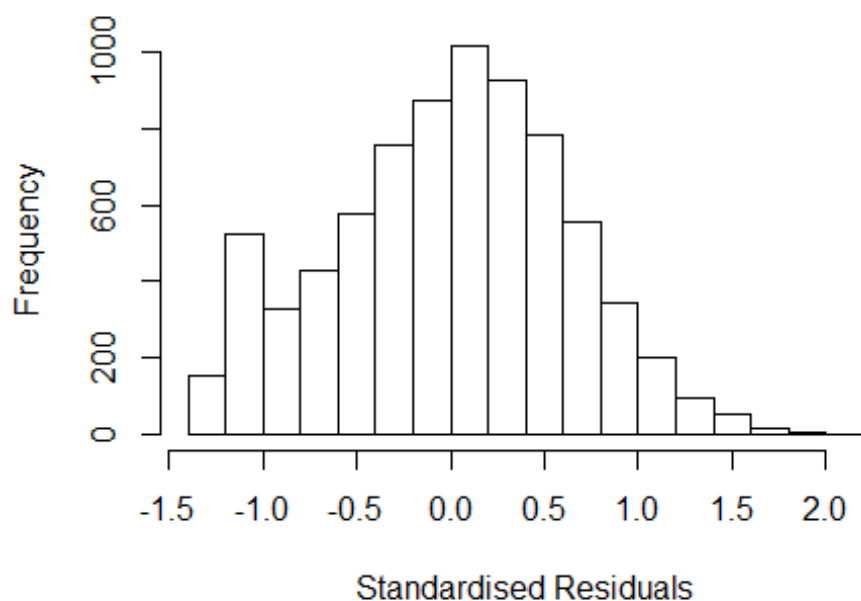
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3740 -0.4274 0.0201 0.4280 2.1105
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9443 0.0338 27.94 < 2e-16 ***
## FirstAuthorFemale1 -0.0147 0.0187 -0.78 0.4339
## LastAuthorFemale1 -0.0161 0.0191 -0.84 0.3997
## UniqueAuthors2 0.2081 0.0170 12.24 < 2e-16 ***
## UniqueAuthors3 0.2549 0.0217 11.75 < 2e-16 ***
## UniqueAuthors4 0.2692 0.0351 7.68 1.9e-14 ***
## UniqueAuthors5 0.2474 0.0547 4.52 6.2e-06 ***
## Year1997 0.0176 0.0501 0.35 0.7262
## Year1998 0.1111 0.0498 2.23 0.0255 *
## Year1999 0.0996 0.0459 2.17 0.0301 *
```

```

## Year2000          0.0833      0.0465      1.79      0.0733 .
## Year2001          0.1425      0.0517      2.76      0.0058 **
## Year2002          0.1155      0.0475      2.43      0.0151 *
## Year2003          0.1133      0.0453      2.50      0.0124 *
## Year2004          0.1454      0.0465      3.13      0.0018 **
## Year2005          0.1748      0.0448      3.90      9.7e-05 ***
## Year2006          0.1056      0.0428      2.47      0.0137 *
## Year2007          0.0909      0.0418      2.17      0.0298 *
## Year2008          0.1367      0.0428      3.20      0.0014 **
## Year2009          0.0707      0.0434      1.63      0.1038
## Year2010          0.1280      0.0437      2.93      0.0034 **
## Year2011          0.1346      0.0443      3.03      0.0024 **
## Year2012          0.1343      0.0427      3.15      0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.629
## Multiple R-squared:  0.0381, Adjusted R-squared:  0.0353
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 651 weights are ~= 1. The remaining 6977 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.237  0.869  0.949  0.910  0.985  0.999
## Algorithmic parameters:
##           tuning.chi              bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.31e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.292 1           1.137
## LastAuthorFemale  1.297 1           1.139
## Year              1.020 16           1.001

```

## Residuals from first and last author



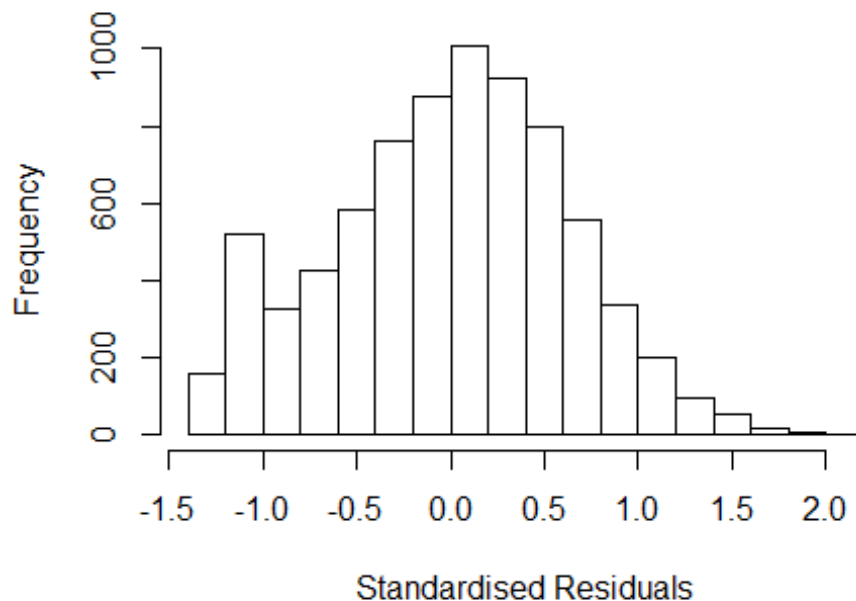
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.237 -0.433  0.035  0.432  2.017
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02592    0.03414   30.05  < 2e-16 ***
## FirstAuthorFemale1 -0.00665    0.01919   -0.35  0.7288
## LastAuthorFemale1 -0.01071    0.01962   -0.55  0.5853
## Year1997         0.02697    0.05179    0.52  0.6025
## Year1998         0.12337    0.05043    2.45  0.0145 *
## Year1999         0.10017    0.04685    2.14  0.0325 *
## Year2000         0.11374    0.04699    2.42  0.0155 *
## Year2001         0.16337    0.05200    3.14  0.0017 **
## Year2002         0.15072    0.04865    3.10  0.0020 **
## Year2003         0.13557    0.04620    2.93  0.0034 **
## Year2004         0.18926    0.04762    3.97  7.1e-05 ***
## Year2005         0.21119    0.04562    4.63  3.7e-06 ***
```

```

## Year2006          0.13180      0.04420      2.98      0.0029 **
## Year2007          0.12533      0.04252      2.95      0.0032 **
## Year2008          0.17501      0.04315      4.06      5.0e-05 ***
## Year2009          0.10496      0.04445      2.36      0.0182 *
## Year2010          0.17587      0.04428      3.97      7.2e-05 ***
## Year2011          0.17768      0.04565      3.89      0.0001 ***
## Year2012          0.18999      0.04342      4.38      1.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.00711,    Adjusted R-squared:  0.00476
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 641 weights are ~= 1. The remaining 6987 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.292  0.864  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.31e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.01 1      1.005
## Year              1.01 16      1.000

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2354 -0.4344 0.0342 0.4305 2.0180
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0249 0.0341 30.06 < 2e-16 ***
## FirstAuthorFemale1 -0.0121 0.0170 -0.71 0.47755
## Year1997 0.0269 0.0518 0.52 0.60275
## Year1998 0.1231 0.0504 2.44 0.01465 *
## Year1999 0.0996 0.0468 2.13 0.03338 *
## Year2000 0.1135 0.0470 2.42 0.01569 *
## Year2001 0.1633 0.0520 3.14 0.00170 **
## Year2002 0.1501 0.0486 3.09 0.00203 **
## Year2003 0.1352 0.0462 2.93 0.00343 **
## Year2004 0.1887 0.0476 3.96 7.4e-05 ***
## Year2005 0.2105 0.0456 4.62 3.9e-06 ***
## Year2006 0.1314 0.0442 2.97 0.00295 **
```

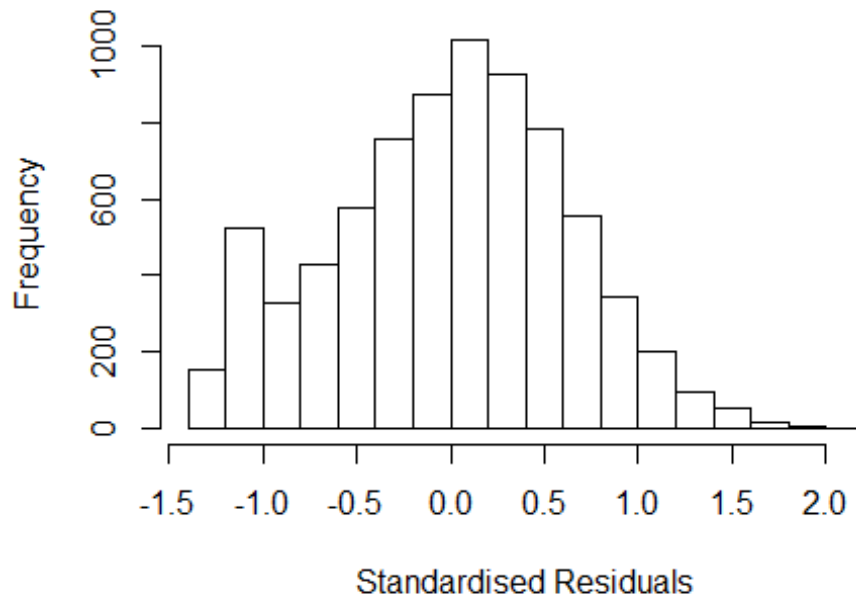
```

## Year2007          0.1249      0.0425      2.94  0.00331 **
## Year2008          0.1745      0.0431      4.04  5.3e-05 ***
## Year2009          0.1045      0.0444      2.35  0.01870 *
## Year2010          0.1756      0.0443      3.97  7.4e-05 ***
## Year2011          0.1769      0.0456      3.88  0.00011 ***
## Year2012          0.1895      0.0434      4.37  1.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.636
## Multiple R-squared:  0.00706,    Adjusted R-squared:  0.00485
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 644 weights are ~= 1. The remaining 6984 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.292  0.864  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.31e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.014 1      1.007
## Year      1.014 16      1.000

```



## Residuals from last author



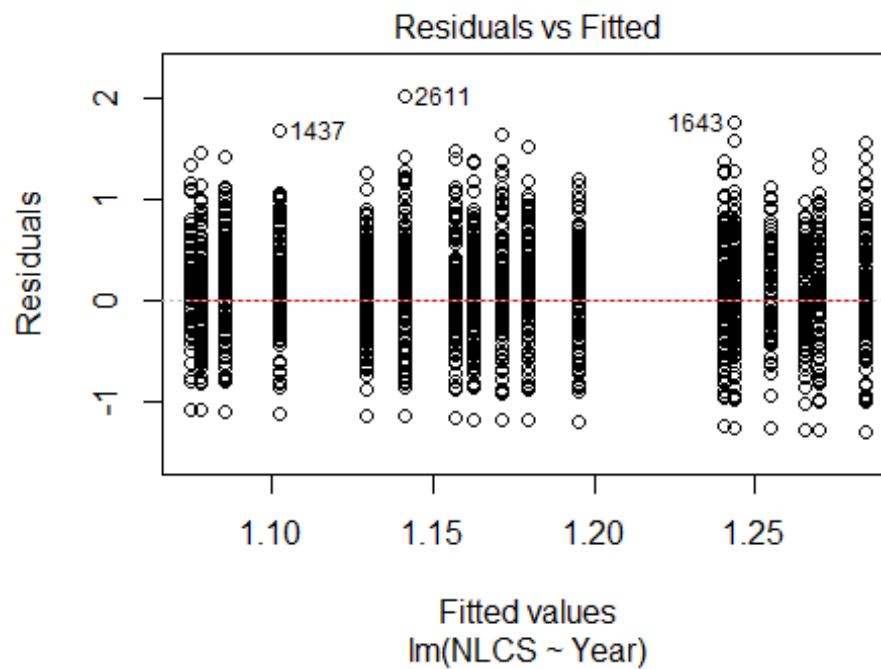
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.236 -0.433 0.034 0.431 2.017
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0252 0.0341 30.09 < 2e-16 ***
## LastAuthorFemale1 -0.0141 0.0174 -0.81 0.4176
## Year1997 0.0270 0.0518 0.52 0.6021
## Year1998 0.1234 0.0504 2.45 0.0144 *
## Year1999 0.1001 0.0469 2.14 0.0326 *
## Year2000 0.1137 0.0470 2.42 0.0156 *
## Year2001 0.1630 0.0520 3.14 0.0017 **
## Year2002 0.1508 0.0487 3.10 0.0019 **
## Year2003 0.1354 0.0462 2.93 0.0034 **
## Year2004 0.1893 0.0476 3.98 7.1e-05 ***
## Year2005 0.2112 0.0456 4.63 3.7e-06 ***
## Year2006 0.1316 0.0442 2.98 0.0029 **
```

```

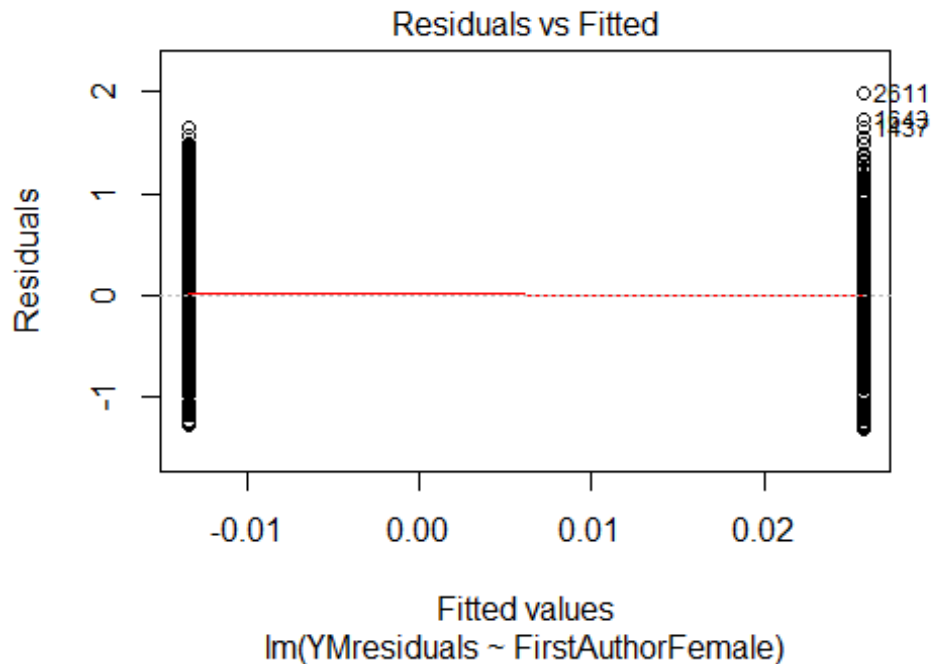
## Year2007          0.1252      0.0425      2.95      0.0032 **
## Year2008          0.1748      0.0431      4.05      5.1e-05 ***
## Year2009          0.1045      0.0444      2.35      0.0186 *
## Year2010          0.1756      0.0443      3.97      7.4e-05 ***
## Year2011          0.1773      0.0456      3.89      0.0001 ***
## Year2012          0.1899      0.0434      4.37      1.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0071, Adjusted R-squared:  0.00488
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 643 weights are ~= 1. The remaining 6985 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.292  0.863  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.31e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7628"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1409"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  116  135  146  151  173  138  117  140  155  162  180  178  168  186  197
## 2011 2012
##   200   211
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   93  114  116  120  143  110   96  119  142  146  156  145  145  152  168
## 2011 2012

```

```
## 175 178
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 91 113 114 115 138 103 90 108 137 135 153 135 136 144 160
## 2011 2012
## 168 166
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 5e-04
```

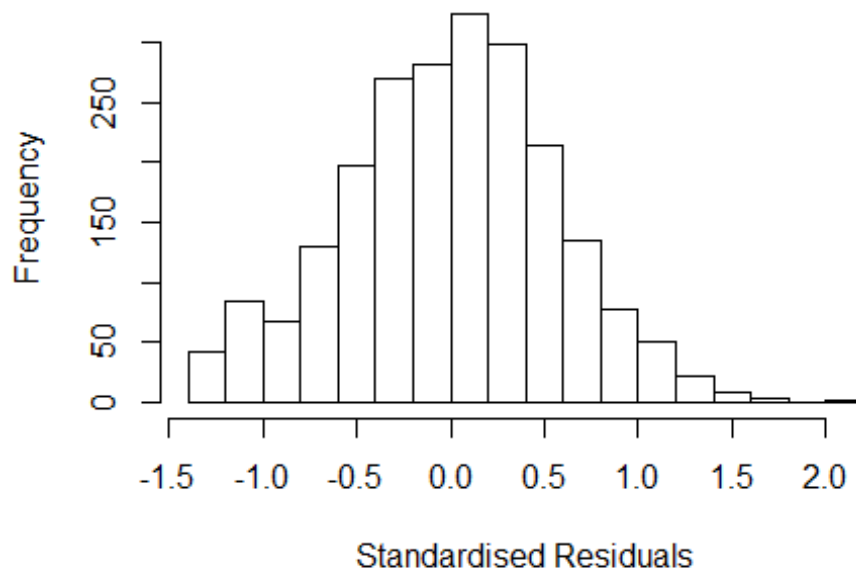


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 1.7609000915299"
## [1] "Male first author team size 2018 geometric mean: 1.66838330007541"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5800, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.58812386612664"
## [1] "Male last author team size 2018 geometric mean: 1.75670933073816"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.291 1          1.136
## LastAuthorFemale  1.281 1          1.132
## UniqueAuthors    1.110 4          1.013
## Year             1.137 16          1.004
```

## Residuals from first and last author and team size



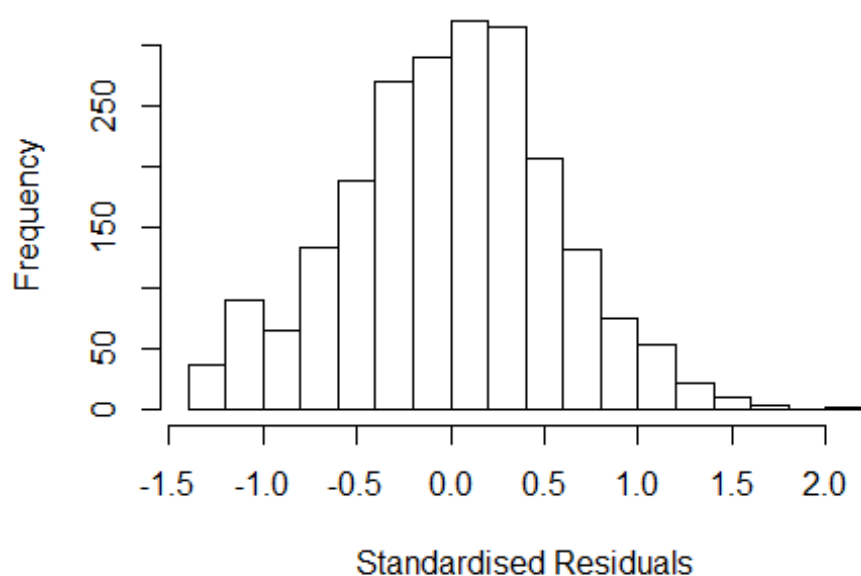
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3728 -0.3665 0.0209 0.3650 2.0001
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04848 0.06932 15.13 <2e-16 ***
## FirstAuthorFemale1 0.03275 0.02899 1.13 0.259
## LastAuthorFemale1 -0.00976 0.02867 -0.34 0.734
## UniqueAuthors2 0.04697 0.02820 1.67 0.096 .
## UniqueAuthors3 0.05395 0.03807 1.42 0.157
## UniqueAuthors4 -0.08415 0.06655 -1.26 0.206
## UniqueAuthors5 0.13564 0.14391 0.94 0.346
## Year1997 0.06955 0.09310 0.75 0.455
## Year1998 0.24459 0.09634 2.54 0.011 *
## Year1999 0.17623 0.08849 1.99 0.047 *
```

```

## Year2000      0.11081    0.08999    1.23    0.218
## Year2001      0.17251    0.09200    1.88    0.061 .
## Year2002      0.19981    0.09175    2.18    0.030 *
## Year2003      0.19107    0.08088    2.36    0.018 *
## Year2004      0.03022    0.07856    0.38    0.701
## Year2005      0.02359    0.08429    0.28    0.780
## Year2006      0.17181    0.08078    2.13    0.034 *
## Year2007     -0.02395    0.08083   -0.30    0.767
## Year2008      0.11067    0.08246    1.34    0.180
## Year2009      0.10129    0.08230    1.23    0.219
## Year2010      0.03547    0.08094    0.44    0.661
## Year2011      0.01579    0.08066    0.20    0.845
## Year2012      0.12575    0.08109    1.55    0.121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.021, Adjusted R-squared:  0.0112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 175 weights are ~= 1. The remaining 2031 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.162  0.872  0.951   0.904   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.299 1      1.140
## LastAuthorFemale  1.281 1      1.132
## Year              1.034 16      1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3441 -0.3677 0.0221 0.3716 2.0284
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06587 0.06898 15.45 <2e-16 ***
## FirstAuthorFemale1 0.03428 0.02918 1.17 0.240
## LastAuthorFemale1 -0.00916 0.02880 -0.32 0.750
## Year1997 0.06886 0.09414 0.73 0.465
## Year1998 0.24399 0.09702 2.51 0.012 *
## Year1999 0.17633 0.08890 1.98 0.047 *
## Year2000 0.11508 0.09013 1.28 0.202
## Year2001 0.17799 0.09225 1.93 0.054 .
## Year2002 0.20142 0.09173 2.20 0.028 *
## Year2003 0.19603 0.08136 2.41 0.016 *
## Year2004 0.03007 0.07920 0.38 0.704
## Year2005 0.02784 0.08456 0.33 0.742
```

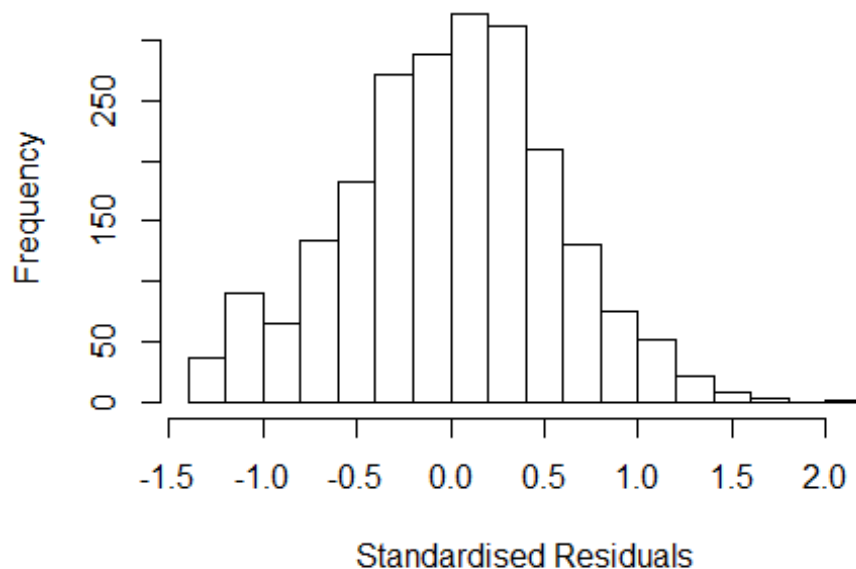
```

## Year2006          0.17404    0.08138    2.14    0.033 *
## Year2007         -0.02123    0.08151   -0.26    0.795
## Year2008          0.11135    0.08308    1.34    0.180
## Year2009          0.10466    0.08302    1.26    0.208
## Year2010          0.03465    0.08153    0.42    0.671
## Year2011          0.01804    0.08134    0.22    0.825
## Year2012          0.12840    0.08155    1.57    0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.0101
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 167 weights are ~= 1. The remaining 2039 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.148  0.870  0.953  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1         1.014
## Year              1.028 16         1.001

```



## Residuals from first author



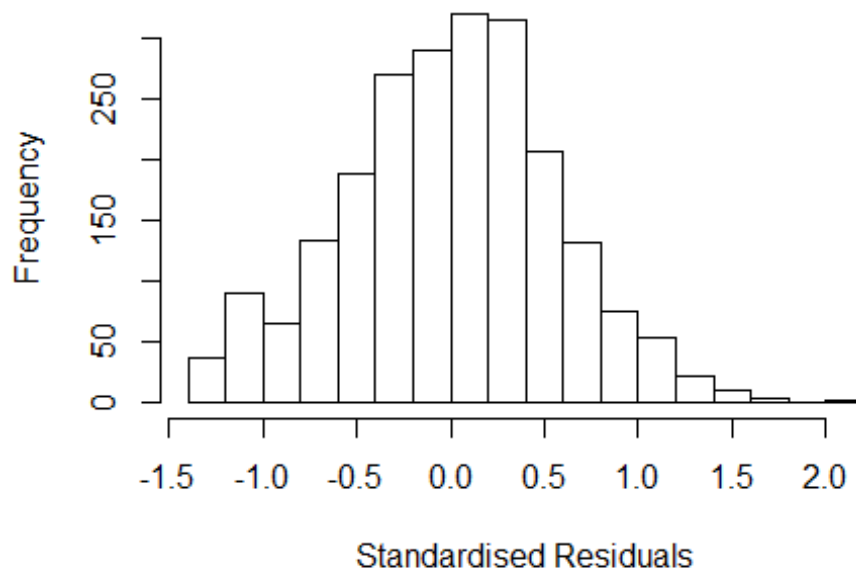
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3381 -0.3662  0.0222  0.3732  2.0258
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0647    0.0688   15.48  <2e-16 ***
## FirstAuthorFemale1 0.0294    0.0260    1.13   0.259
## Year1997          0.0685    0.0942    0.73   0.467
## Year1998          0.2441    0.0971    2.51   0.012 *
## Year1999          0.1761    0.0890    1.98   0.048 *
## Year2000          0.1149    0.0902    1.27   0.203
## Year2001          0.1779    0.0923    1.93   0.054 .
## Year2002          0.2011    0.0918    2.19   0.028 *
## Year2003          0.1957    0.0815    2.40   0.016 *
## Year2004          0.0295    0.0793    0.37   0.710
## Year2005          0.0276    0.0846    0.33   0.745
## Year2006          0.1735    0.0814    2.13   0.033 *
```

```

## Year2007          -0.0216      0.0816   -0.27    0.791
## Year2008          0.1115      0.0831    1.34    0.180
## Year2009          0.1048      0.0831    1.26    0.207
## Year2010          0.0342      0.0816    0.42    0.675
## Year2011          0.0177      0.0814    0.22    0.828
## Year2012          0.1277      0.0816    1.56    0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0181, Adjusted R-squared:  0.0105
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 2037 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.150  0.869  0.953  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.015 1          1.007
## Year              1.015 16          1.000

```

## Residuals from last author



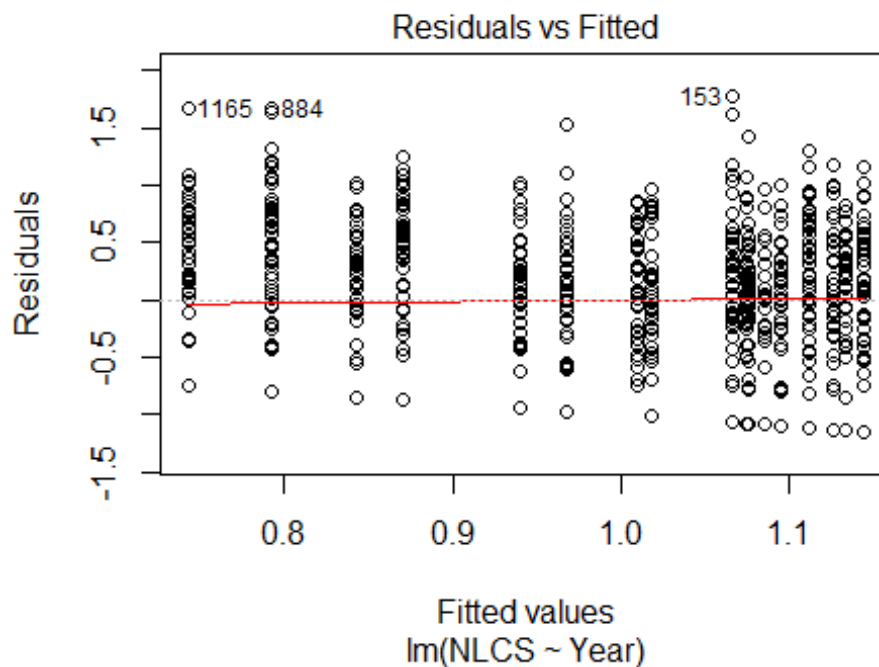
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3233 -0.3638 0.0217 0.3709 2.0382
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07197 0.06894 15.55 <2e-16 ***
## LastAuthorFemale1 0.00897 0.02570 0.35 0.727
## Year1997 0.06582 0.09408 0.70 0.484
## Year1998 0.24239 0.09708 2.50 0.013 *
## Year1999 0.17515 0.08899 1.97 0.049 *
## Year2000 0.11182 0.09034 1.24 0.216
## Year2001 0.17659 0.09232 1.91 0.056 .
## Year2002 0.20081 0.09164 2.19 0.029 *
## Year2003 0.19488 0.08151 2.39 0.017 *
## Year2004 0.02934 0.07926 0.37 0.711
## Year2005 0.02623 0.08456 0.31 0.756
## Year2006 0.17259 0.08143 2.12 0.034 *
```

```

## Year2007          -0.02090      0.08155    -0.26      0.798
## Year2008           0.11383      0.08323      1.37      0.172
## Year2009           0.10546      0.08313      1.27      0.205
## Year2010           0.03484      0.08166      0.43      0.670
## Year2011           0.01910      0.08150      0.23      0.815
## Year2012           0.12818      0.08167      1.57      0.117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0176, Adjusted R-squared:  0.00999
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 179 weights are ~= 1. The remaining 2027 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.143  0.867  0.952   0.903  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2206"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1410"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   77   73   56   69   50   51   54   65   59   38   35   67   52  141  116
## 2011 2012
##   91   53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   56   55   43   54   42   38   38   55   47   31   26   48   41  121   98
## 2011 2012

```

```
## 78 49
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 54 40 53 42 35 37 50 44 30 24 41 37 115 92
## 2011 2012
## 72 47
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



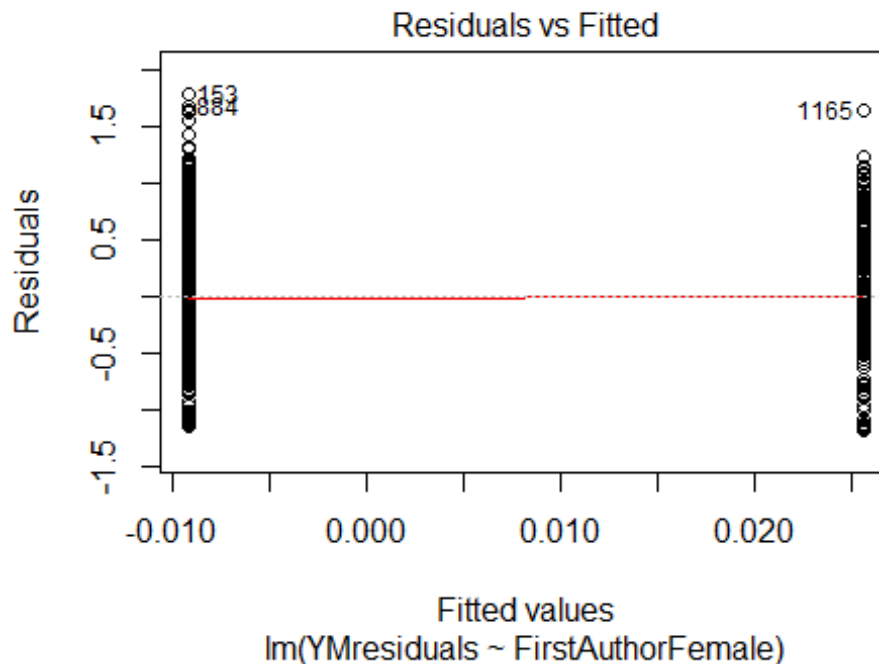
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.03, df = 1, p-value = 0.9

## [1] "Female first author team size 2018 geometric mean: 1.64375182951723"
## [1] "Male first author team size 2018 geometric mean: 1.44317687342364"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

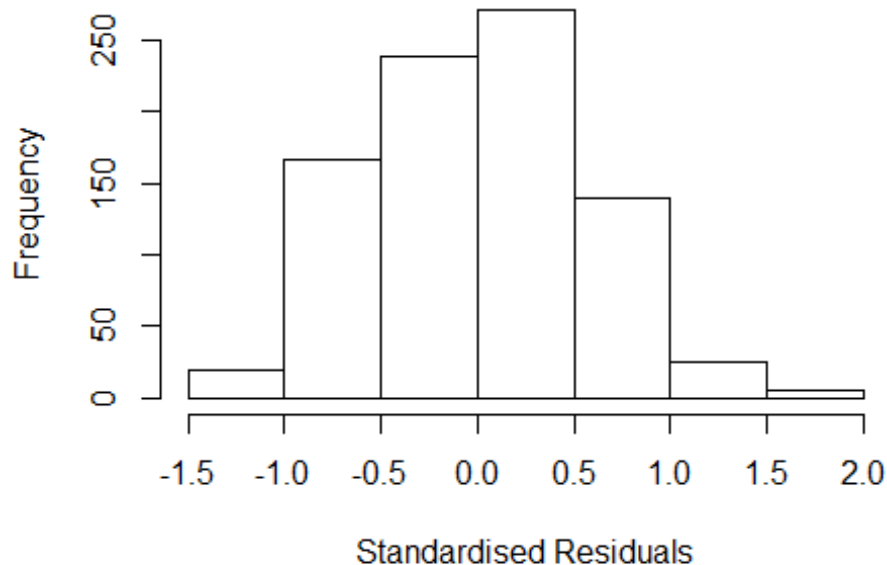
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.64375182951723"
## [1] "Male last author team size 2018 geometric mean: 1.44317687342364"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.677  1      1.295
## LastAuthorFemale  1.658  1      1.287
## UniqueAuthors    1.317  4      1.035
## Year              1.455 16      1.012
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3115 -0.3962 0.0155 0.3943 1.8191
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8446 0.0724 11.67 < 2e-16 ***
## FirstAuthorFemale1 0.0528 0.0580 0.91 0.36316
## LastAuthorFemale1 -0.0592 0.0592 -1.00 0.31705
## UniqueAuthors2 0.2559 0.0479 5.35 1.1e-07 ***
## UniqueAuthors3 0.2994 0.0634 4.72 2.7e-06 ***
## UniqueAuthors4 0.2967 0.1174 2.53 0.01170 *
## UniqueAuthors5 0.5166 0.1372 3.76 0.00018 ***
## Year1997 0.1229 0.1192 1.03 0.30282
## Year1998 0.0443 0.1003 0.44 0.65895
## Year1999 -0.0946 0.1006 -0.94 0.34705
```

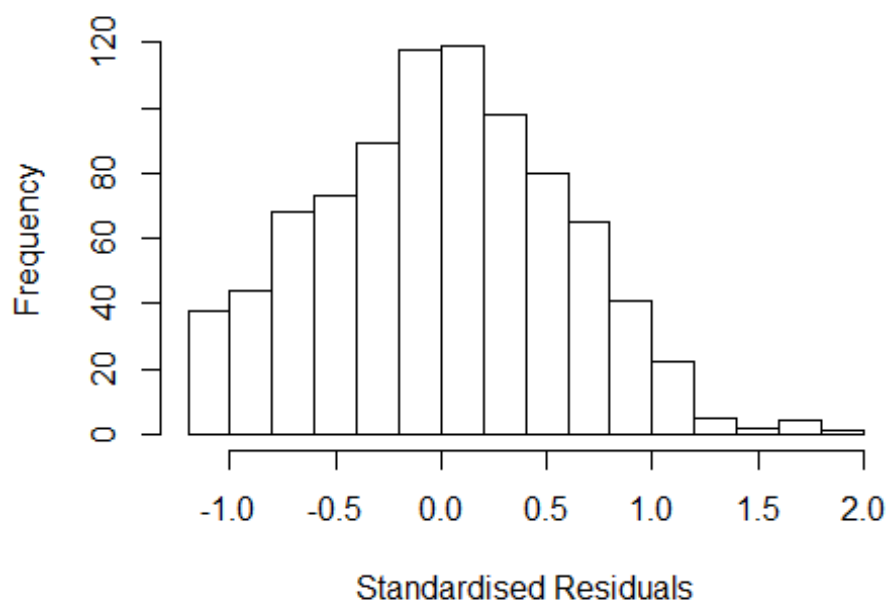
```

## Year2000          0.0460      0.0994      0.46  0.64345
## Year2001          0.1861      0.1033      1.80  0.07197 .
## Year2002          0.1131      0.0975      1.16  0.24652
## Year2003          0.1676      0.1291      1.30  0.19467
## Year2004          0.1525      0.1029      1.48  0.13866
## Year2005          0.0942      0.1187      0.79  0.42747
## Year2006          0.1293      0.1105      1.17  0.24203
## Year2007          0.1282      0.1185      1.08  0.27938
## Year2008          0.0792      0.1237      0.64  0.52227
## Year2009         -0.2197      0.0933     -2.36  0.01874 *
## Year2010         -0.1357      0.1015     -1.34  0.18183
## Year2011         -0.2402      0.1078     -2.23  0.02621 *
## Year2012          0.0159      0.1061      0.15  0.88104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0882
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 786 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.300  0.873  0.944  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.650 1      1.285
## LastAuthorFemale  1.661 1      1.289
## Year              1.140 16      1.004

```



## Residuals from first and last author



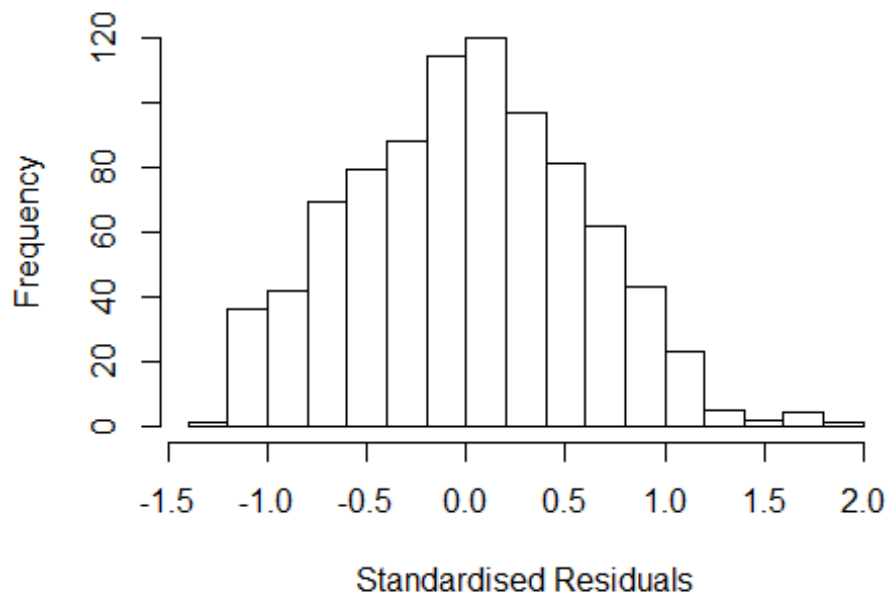
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.17599 -0.41626 0.00402 0.40203 1.86973
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9513 0.0698 13.63 <2e-16 ***
## FirstAuthorFemale1 0.0912 0.0599 1.52 0.128
## LastAuthorFemale1 -0.0720 0.0616 -1.17 0.243
## Year1997 0.0890 0.1189 0.75 0.454
## Year1998 0.0508 0.1038 0.49 0.625
## Year1999 -0.0963 0.1054 -0.91 0.361
## Year2000 0.0374 0.1031 0.36 0.717
## Year2001 0.2055 0.1067 1.93 0.054 .
## Year2002 0.1423 0.1009 1.41 0.159
## Year2003 0.1501 0.1290 1.16 0.245
## Year2004 0.1731 0.1056 1.64 0.101
## Year2005 0.0962 0.1218 0.79 0.430
```

```

## Year2006          0.1028      0.1185      0.87      0.386
## Year2007          0.1582      0.1262      1.25      0.210
## Year2008          0.1053      0.1237      0.85      0.395
## Year2009         -0.2143      0.0958     -2.24      0.026 *
## Year2010         -0.1196      0.1007     -1.19      0.235
## Year2011         -0.2300      0.1082     -2.13      0.034 *
## Year2012          0.0430      0.1087      0.40      0.692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0596, Adjusted R-squared:  0.0396
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 800 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.310  0.867  0.950  0.913  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.064 1          1.032
## Year              1.064 16          1.002

```

## Residuals from first author



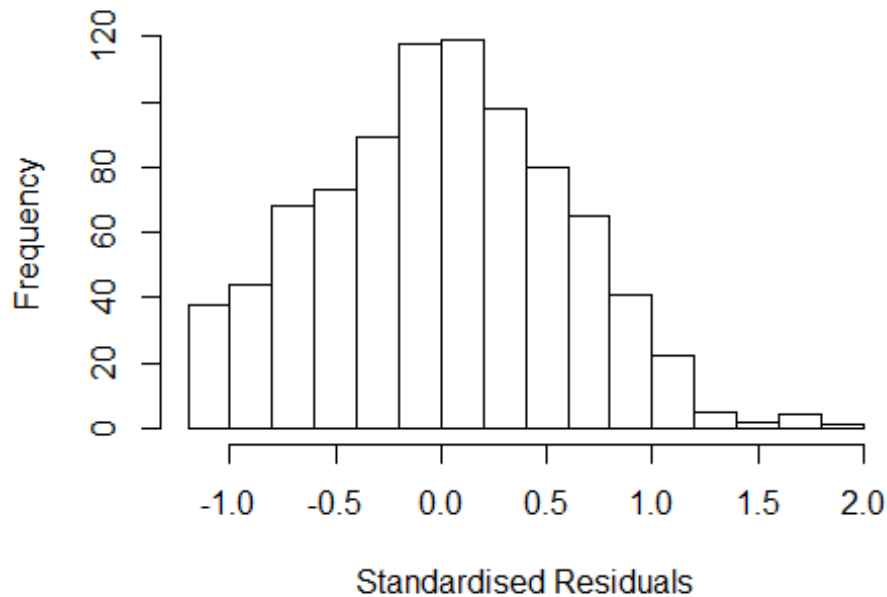
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20035 -0.41690  0.00799  0.40567  1.80457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9407    0.0686   13.72  <2e-16 ***
## FirstAuthorFemale1 0.0483    0.0485    0.99   0.320
## Year1997          0.0927    0.1192    0.78   0.437
## Year1998          0.0510    0.1036    0.49   0.623
## Year1999         -0.0897    0.1044   -0.86   0.391
## Year2000          0.0446    0.1025    0.43   0.664
## Year2001          0.2114    0.1061    1.99   0.047 *
## Year2002          0.1499    0.1006    1.49   0.137
## Year2003          0.1595    0.1282    1.24   0.214
## Year2004          0.1758    0.1052    1.67   0.095 .
## Year2005          0.1012    0.1220    0.83   0.407
## Year2006          0.1071    0.1184    0.90   0.366
```

```

## Year2007          0.1519      0.1235      1.23      0.219
## Year2008          0.1137      0.1237      0.92      0.358
## Year2009         -0.2070      0.0953     -2.17      0.030 *
## Year2010         -0.1170      0.0998     -1.17      0.242
## Year2011         -0.2280      0.1083     -2.11      0.036 *
## Year2012          0.0467      0.1083      0.43      0.666
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.0578, Adjusted R-squared:  0.039
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 799 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.347  0.869  0.952  0.914  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.068 1      1.033
## Year              1.068 16      1.002

```

## Residuals from last author



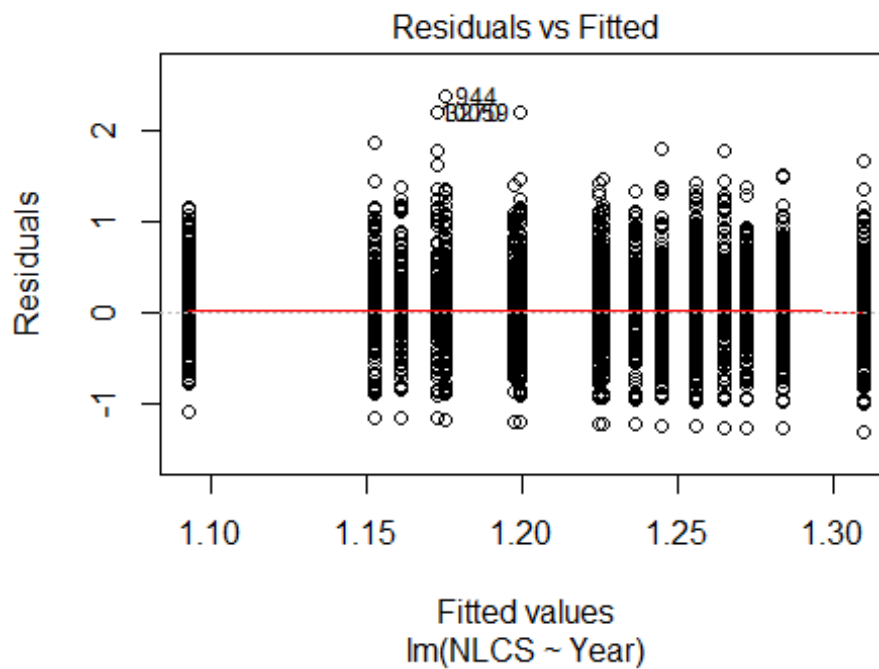
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15906 -0.40318 0.00779 0.40879 1.80631
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9530 0.0692 13.77 <2e-16 ***
## LastAuthorFemale1 -0.0158 0.0491 -0.32 0.748
## Year1997 0.0945 0.1191 0.79 0.428
## Year1998 0.0500 0.1038 0.48 0.630
## Year1999 -0.0884 0.1036 -0.85 0.394
## Year2000 0.0542 0.1009 0.54 0.591
## Year2001 0.2219 0.1063 2.09 0.037 *
## Year2002 0.1516 0.1010 1.50 0.134
## Year2003 0.1674 0.1275 1.31 0.190
## Year2004 0.1819 0.1050 1.73 0.084 .
## Year2005 0.1112 0.1215 0.91 0.361
## Year2006 0.1017 0.1176 0.86 0.388
```

```

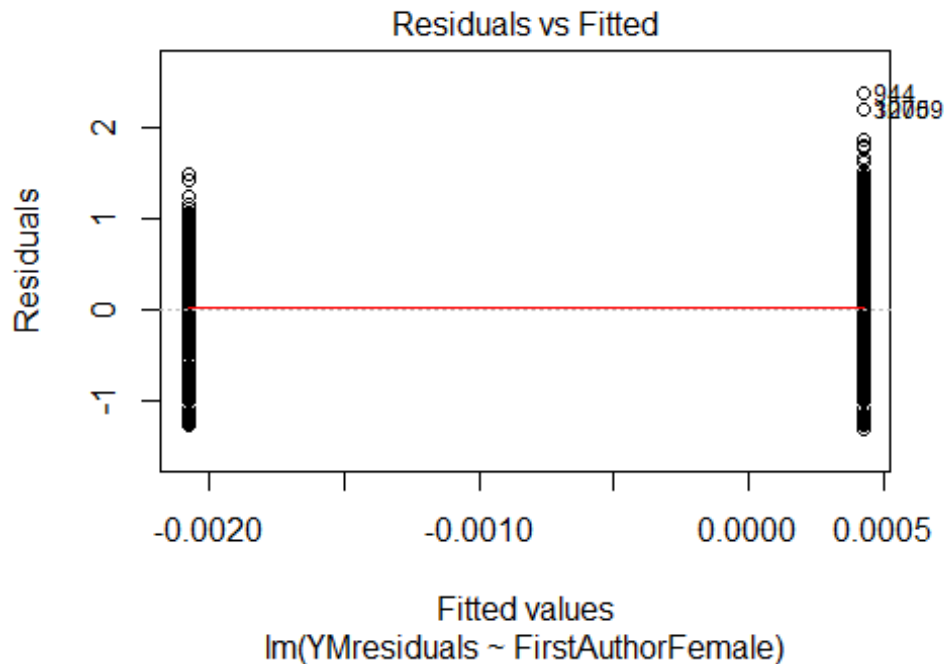
## Year2007          0.1616      0.1251      1.29      0.197
## Year2008          0.1221      0.1243      0.98      0.326
## Year2009         -0.2028      0.0953     -2.13      0.034 *
## Year2010         -0.1125      0.1001     -1.12      0.261
## Year2011         -0.2263      0.1081     -2.09      0.037 *
## Year2012          0.0490      0.1087      0.45      0.652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0568, Adjusted R-squared:  0.038
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 796 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.342  0.863  0.949   0.913  0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.15e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 867"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1500"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 630 676 593 526 601 575 564 483 481 480 502 468 441 561 568
## 2011 2012
## 583 581
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 166 198 168 143 135 131 176 148 149 167 173 168 189 243 264
## 2011 2012

```

```
## 276 294
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 134 174 150 126 111 112 142 124 123 136 137 133 154 199 219
## 2011 2012
## 231 253
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 53, df = 16, p-value = 8e-06
```



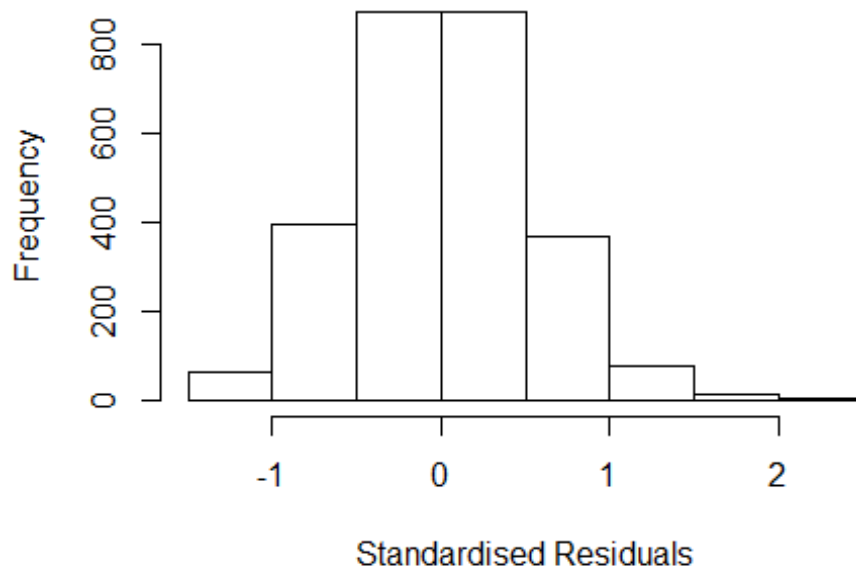
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.8, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 4.19470563277496"
## [1] "Male first author team size 2018 geometric mean: 3.2199227919911"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9400, p-value = 3e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.63832419766093"
## [1] "Male last author team size 2018 geometric mean: 3.43504932601334"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5500, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1          1.031
## LastAuthorFemale 1.051 1          1.025
## UniqueAuthors    1.131 4          1.016
## Year              1.158 16         1.005
```



## Residuals from first and last author and team size



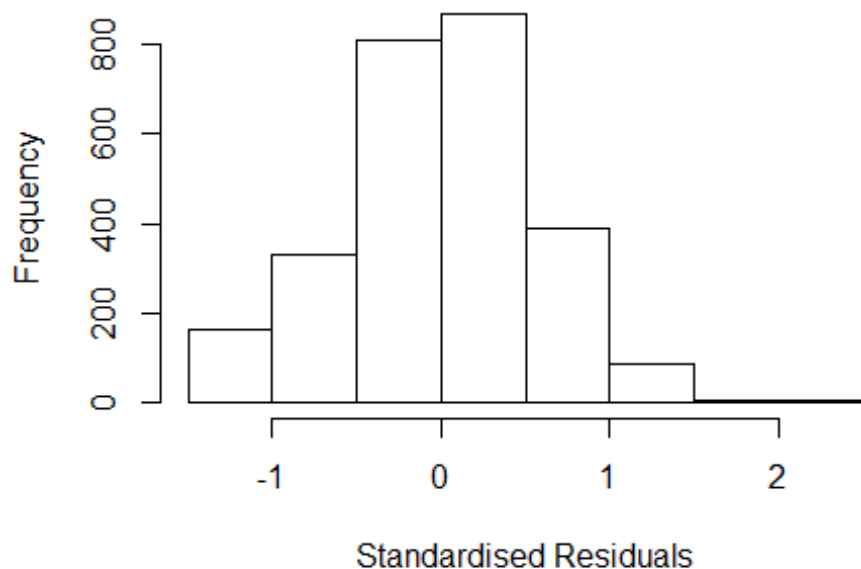
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.423929 -0.350900 0.000715 0.354166 2.274200
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9202 0.0610 15.10 < 2e-16 ***
## FirstAuthorFemale1 -0.0518 0.0269 -1.92 0.05484 .
## LastAuthorFemale1 -0.0668 0.0307 -2.18 0.02964 *
## UniqueAuthors2 0.4418 0.0398 11.10 < 2e-16 ***
## UniqueAuthors3 0.4552 0.0402 11.32 < 2e-16 ***
## UniqueAuthors4 0.5178 0.0440 11.76 < 2e-16 ***
## UniqueAuthors5 0.5556 0.0412 13.48 < 2e-16 ***
## Year1997 -0.0927 0.0693 -1.34 0.18077
## Year1998 -0.0240 0.0718 -0.33 0.73778
## Year1999 -0.0519 0.0768 -0.68 0.49889
```

```

## Year2000          -0.1094      0.0835   -1.31   0.19033
## Year2001          -0.0886      0.0763   -1.16   0.24566
## Year2002          -0.0303      0.0679   -0.45   0.65511
## Year2003          -0.0859      0.0727   -1.18   0.23783
## Year2004          -0.1136      0.0715   -1.59   0.11217
## Year2005          -0.1683      0.0715   -2.35   0.01870 *
## Year2006          -0.1754      0.0733   -2.39   0.01671 *
## Year2007          -0.0567      0.0743   -0.76   0.44569
## Year2008          -0.0967      0.0673   -1.44   0.15072
## Year2009          -0.0741      0.0663   -1.12   0.26386
## Year2010          -0.0446      0.0611   -0.73   0.46489
## Year2011          -0.1581      0.0611   -2.59   0.00977 **
## Year2012          -0.2318      0.0617   -3.76   0.00018 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.529
## Multiple R-squared:  0.118, Adjusted R-squared:  0.11
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.025  0.858  0.952  0.903  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.76e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1          1.015
## LastAuthorFemale  1.036 1          1.018
## Year              1.051 16          1.002

```

## Residuals from first and last author



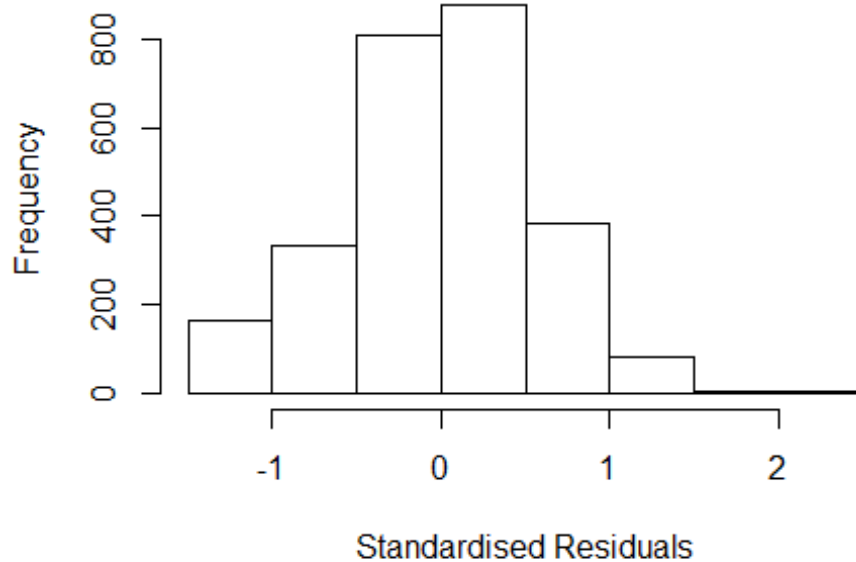
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2947 -0.3497 0.0161 0.3669 2.3541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28942 0.05700 22.62 <2e-16 ***
## FirstAuthorFemale1 -0.00887 0.02791 -0.32 0.7506
## LastAuthorFemale1 -0.05525 0.03215 -1.72 0.0858 .
## Year1997 -0.10149 0.07357 -1.38 0.1679
## Year1998 -0.00887 0.07562 -0.12 0.9067
## Year1999 -0.07954 0.08374 -0.95 0.3423
## Year2000 -0.12926 0.09572 -1.35 0.1770
## Year2001 -0.11182 0.08051 -1.39 0.1650
## Year2002 0.00518 0.07279 0.07 0.9433
## Year2003 -0.05278 0.08101 -0.65 0.5148
## Year2004 -0.06928 0.07510 -0.92 0.3564
## Year2005 -0.08849 0.07602 -1.16 0.2446
```

```

## Year2006          -0.12451      0.07708      -1.62      0.1063
## Year2007          -0.01184      0.07724      -0.15      0.8782
## Year2008          -0.03942      0.07220      -0.55      0.5851
## Year2009          -0.04851      0.07049      -0.69      0.4914
## Year2010           0.00530      0.06564       0.08      0.9357
## Year2011          -0.11602      0.06548      -1.77      0.0765 .
## Year2012          -0.17153      0.06606      -2.60      0.0095 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.00419
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 238 weights are ~= 1. The remaining 2420 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0185 0.8600 0.9510 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.76e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year              1.025 16      1.001

```

## Residuals from first author



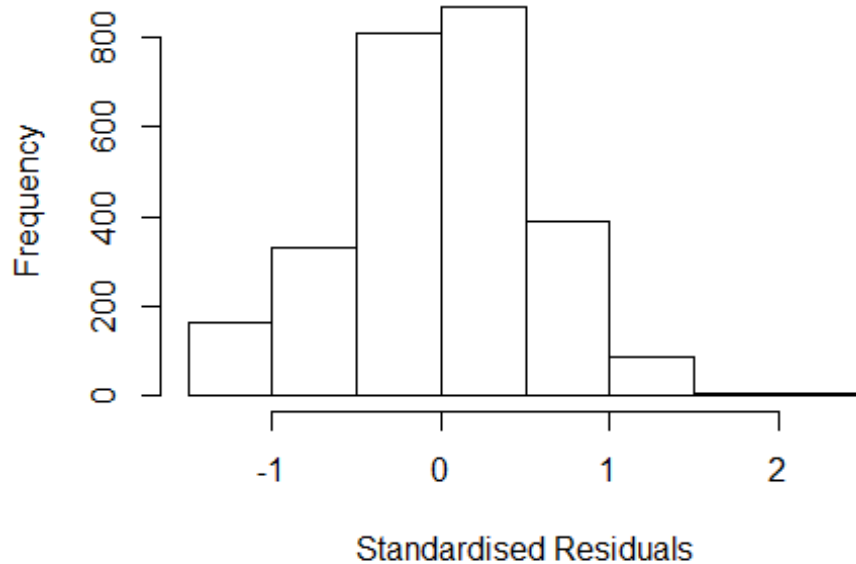
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2902 -0.3482 0.0109 0.3717 2.3585
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28445 0.05696 22.55 <2e-16 ***
## FirstAuthorFemale1 -0.01691 0.02808 -0.60 0.5472
## Year1997 -0.10096 0.07353 -1.37 0.1699
## Year1998 -0.00606 0.07557 -0.08 0.9361
## Year1999 -0.07966 0.08374 -0.95 0.3416
## Year2000 -0.13095 0.09569 -1.37 0.1713
## Year2001 -0.11382 0.08076 -1.41 0.1588
## Year2002 0.00440 0.07298 0.06 0.9519
## Year2003 -0.05470 0.08089 -0.68 0.4990
## Year2004 -0.06798 0.07503 -0.91 0.3650
## Year2005 -0.08846 0.07610 -1.16 0.2451
## Year2006 -0.12345 0.07722 -1.60 0.1100
```

```

## Year2007          -0.01385    0.07727   -0.18    0.8578
## Year2008          -0.04037    0.07233   -0.56    0.5768
## Year2009          -0.05224    0.07047   -0.74    0.4586
## Year2010           0.00572    0.06572    0.09    0.9307
## Year2011          -0.11556    0.06557   -1.76    0.0781 .
## Year2012          -0.17211    0.06608   -2.60    0.0092 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.00999,    Adjusted R-squared:  0.00361
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 222 weights are ~= 1. The remaining 2436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0185 0.8620 0.9520 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.76e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.015
## Year            1.029 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2931 -0.3506 0.0145 0.3660 2.3550
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28820 0.05688 22.65 <2e-16 ***
## LastAuthorFemale1 -0.05704 0.03220 -1.77 0.0766 .
## Year1997 -0.10124 0.07357 -1.38 0.1689
## Year1998 -0.00866 0.07570 -0.11 0.9089
## Year1999 -0.07944 0.08380 -0.95 0.3432
## Year2000 -0.12835 0.09569 -1.34 0.1799
## Year2001 -0.11196 0.08058 -1.39 0.1648
## Year2002 0.00453 0.07276 0.06 0.9504
## Year2003 -0.05221 0.08106 -0.64 0.5196
## Year2004 -0.06912 0.07512 -0.92 0.3576
## Year2005 -0.08864 0.07603 -1.17 0.2438
## Year2006 -0.12485 0.07713 -1.62 0.1057
```

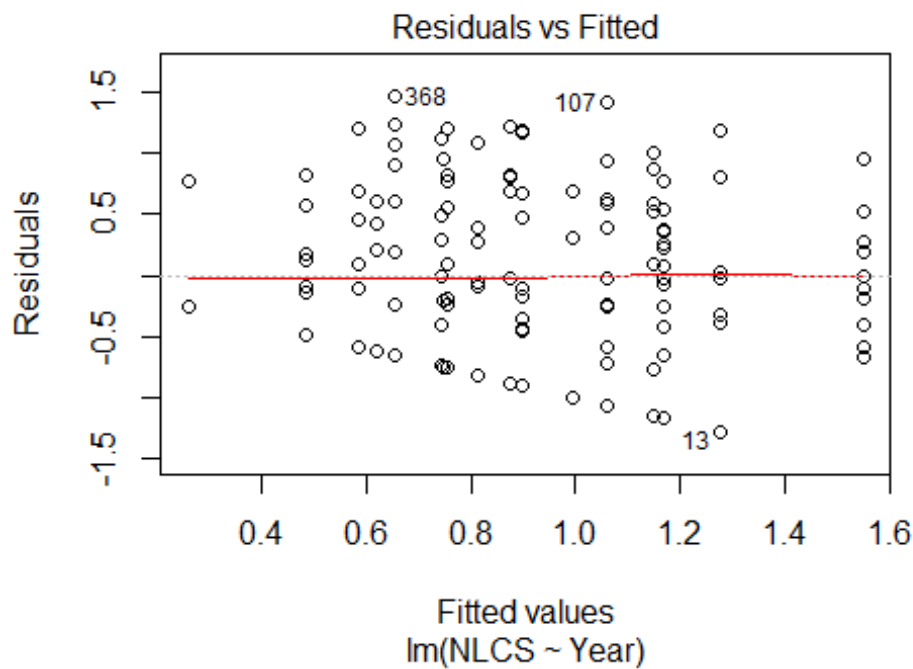
```

## Year2007          -0.01188      0.07727    -0.15    0.8779
## Year2008          -0.04008      0.07220    -0.56    0.5788
## Year2009          -0.04843      0.07051    -0.69    0.4922
## Year2010           0.00489      0.06568     0.07    0.9406
## Year2011          -0.11660      0.06554    -1.78    0.0753 .
## Year2012          -0.17227      0.06602    -2.61    0.0091 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.00452
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 239 weights are ~= 1. The remaining 2419 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0181 0.8610 0.9510 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.76e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2658"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1501"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   54   22   32   37   23   31   33   14   18   15   22   23    9    8
## 2011 2012
##   10   14
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   14   12    9    8    5    9    9    7   10   11   15   13    3    3
## 2011 2012

```



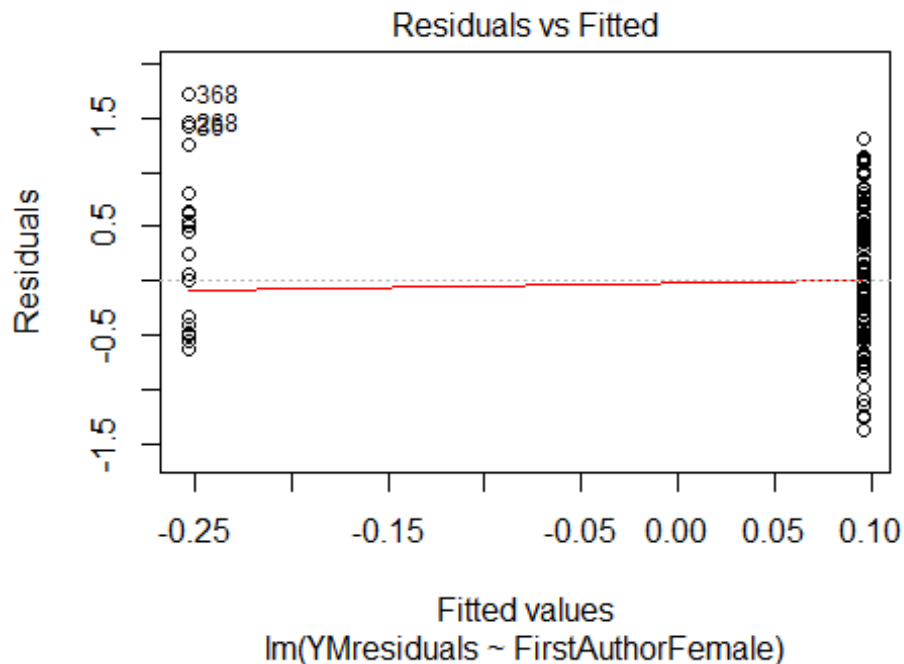
```
##      4      7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      6     11     12      9      7      4      8      9      6      9      9     14     12      2      3
## 2011 2012
##      4      6
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 9.5, df = 16, p-value = 0.9
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.092, df = 1, p-value = 0.8
## [1] "Female first author team size 2018 geometric mean: 2.40462687171946"
## [1] "Male first author team size 2018 geometric mean: 3.27516199844705"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 68, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.56508458007329"
## [1] "Male last author team size 2018 geometric mean: 3.29884978156268"

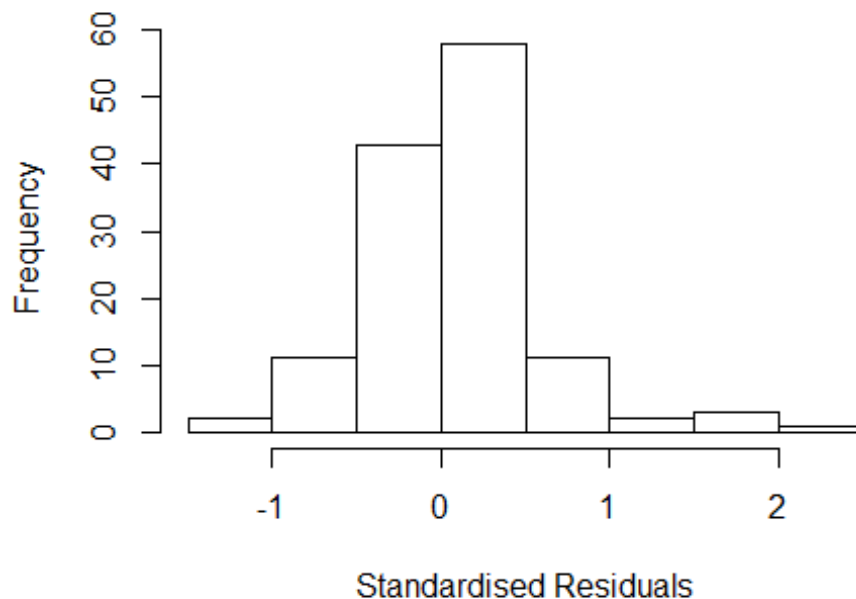
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	22.41	1	4.734
LastAuthorFemale	17.55	1	4.189
UniqueAuthors	468.62	4	2.157
Year	1463.37	16	1.256

## Residuals from first and last author and team size



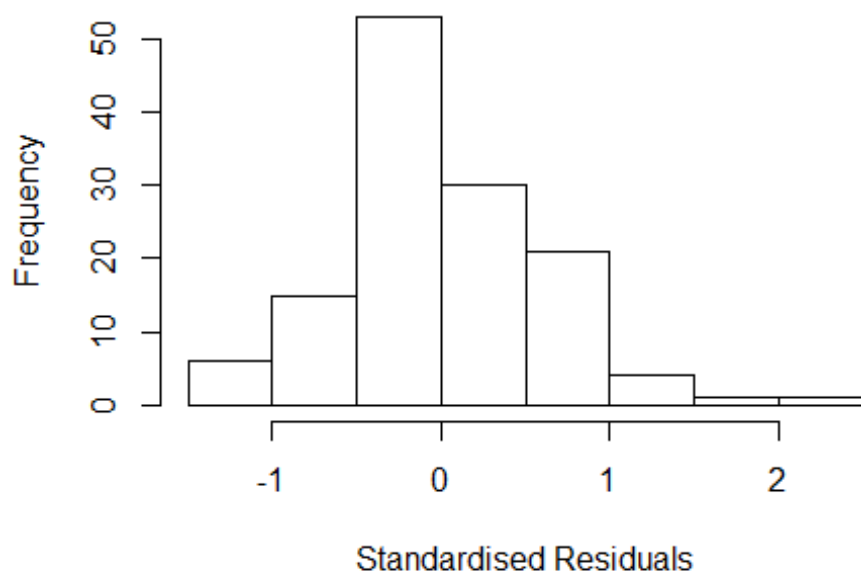
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4886 -0.1856 0.0425 0.2146 2.1615
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7816 0.4529 1.73 0.08724 .
## FirstAuthorFemale1 -0.4359 0.1849 -2.36 0.02022 *
## LastAuthorFemale1 -0.2168 0.2009 -1.08 0.28284
## UniqueAuthors2 0.1001 0.1858 0.54 0.59099
## UniqueAuthors3 0.8524 0.1911 4.46 2.0e-05 ***
## UniqueAuthors4 1.0385 0.2254 4.61 1.1e-05 ***
## UniqueAuthors5 0.9683 0.2602 3.72 0.00032 ***
## Year1997 -0.3549 0.4768 -0.74 0.45822
## Year1998 -0.0238 0.7296 -0.03 0.97401
## Year1999 -0.4427 0.4953 -0.89 0.37341
```

```

## Year2000          -0.7147      0.5169    -1.38   0.16959
## Year2001          -0.1455      0.9380    -0.16   0.87702
## Year2002           0.0567      0.4787     0.12   0.90590
## Year2003           0.0300      0.4735     0.06   0.94962
## Year2004          -0.1795      0.4952    -0.36   0.71770
## Year2005           0.1263      0.5544     0.23   0.82027
## Year2006          -0.1829      0.4904    -0.37   0.70995
## Year2007          -0.1714      0.4770    -0.36   0.72001
## Year2008           0.0724      0.5073     0.14   0.88678
## Year2009          -0.5916      0.5206    -1.14   0.25834
## Year2010          -0.1607      0.5504    -0.29   0.77079
## Year2011          -0.2392      0.4915    -0.49   0.62747
## Year2012          -0.1198      0.5038    -0.24   0.81248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.357
## Multiple R-squared:  0.715, Adjusted R-squared:  0.657
## Convergence in 39 IRWLS iterations
##
## Robustness weights:
## 3 observations c(6,41,104) are outliers with |weight| = 0 ( < 0.00076);
## 10 weights are ~= 1. The remaining 118 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0222 0.8680 0.9660 0.8720 0.9910 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           7.63e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 16.896 1 4.110
## LastAuthorFemale 12.278 1 3.504
## Year 7.073 16 1.063

```

## Residuals from first and last author



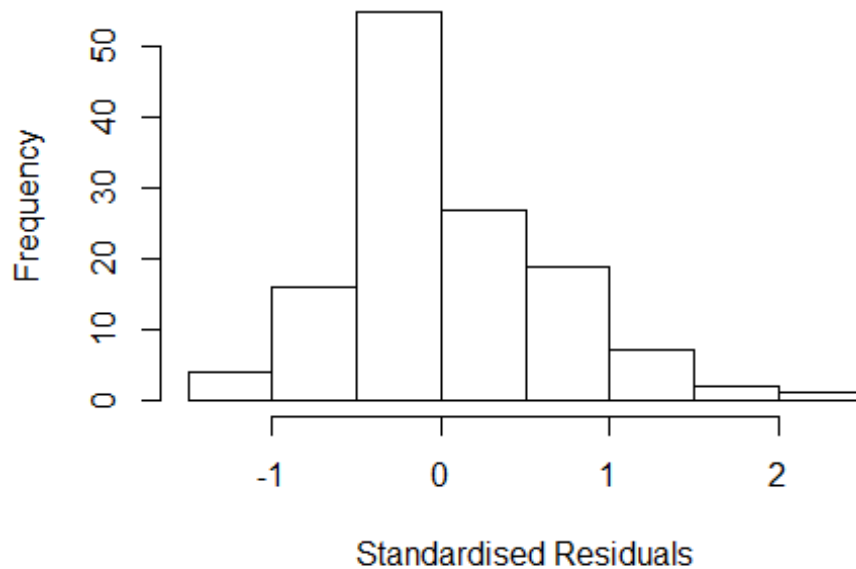
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.255 -0.350 -0.082 0.429 2.074
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.441 0.276 5.21 8.6e-07 ***
## FirstAuthorFemale1 -0.362 0.304 -1.19 0.2361
## LastAuthorFemale1 -0.632 0.251 -2.52 0.0131 *
## Year1997 -0.805 0.347 -2.32 0.0223 *
## Year1998 -0.284 0.419 -0.68 0.4993
## Year1999 -0.970 0.316 -3.07 0.0027 **
## Year2000 -0.360 0.771 -0.47 0.6415
## Year2001 -0.977 0.386 -2.53 0.0128 *
## Year2002 -0.191 0.335 -0.57 0.5693
## Year2003 -0.410 0.319 -1.29 0.2012
## Year2004 -0.546 0.313 -1.74 0.0839 .
## Year2005 0.201 0.339 0.59 0.5539
```

```

## Year2006          -0.268      0.382   -0.70   0.4838
## Year2007          -0.325      0.322   -1.01   0.3156
## Year2008          -0.186      0.336   -0.55   0.5811
## Year2009          -0.474      0.393   -1.20   0.2309
## Year2010          -0.116      0.588   -0.20   0.8441
## Year2011          -0.437      0.304   -1.44   0.1538
## Year2012          -0.263      0.365   -0.72   0.4733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.43,   Adjusted R-squared:  0.338
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 121 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.114  0.863  0.942  0.893  0.990  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.63e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 8.157 1      2.856
## Year              8.157 16      1.068

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2236 -0.3023 -0.0816 0.4374 2.1232
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4409 0.2763 5.22 8.4e-07 ***
## FirstAuthorFemale1 -0.8896 0.2397 -3.71 0.00032 ***
## Year1997 -0.8289 0.3437 -2.41 0.01747 *
## Year1998 -0.3673 0.4523 -0.81 0.41845
## Year1999 -0.9698 0.3158 -3.07 0.00267 **
## Year2000 -0.5305 0.8253 -0.64 0.52165
## Year2001 -0.9768 0.3862 -2.53 0.01281 *
## Year2002 -0.2493 0.3640 -0.68 0.49480
## Year2003 -0.4946 0.3185 -1.55 0.12322
## Year2004 -0.5836 0.3259 -1.79 0.07606 .
## Year2005 0.1661 0.3533 0.47 0.63909
## Year2006 -0.2746 0.4207 -0.65 0.51521
```

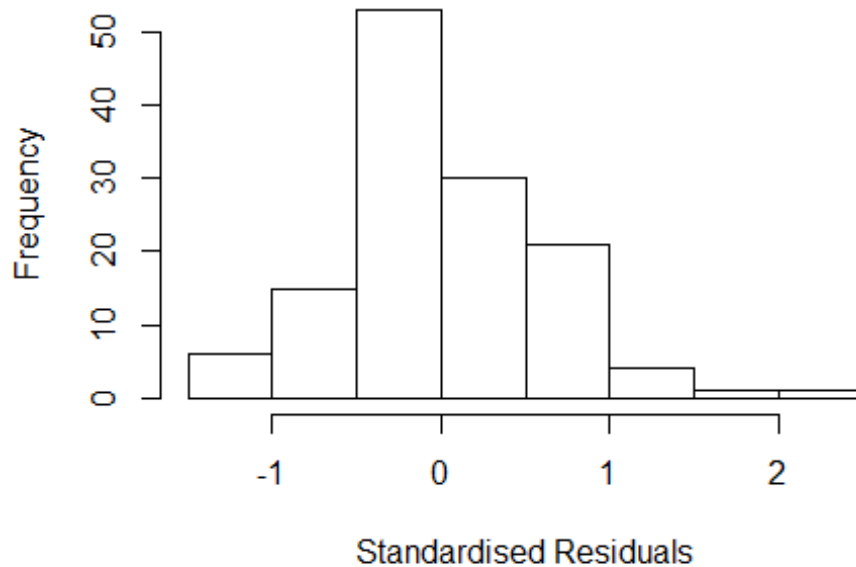
```

## Year2007          -0.3508      0.3584   -0.98   0.32975
## Year2008          -0.2173      0.3328   -0.65   0.51507
## Year2009          -0.7899      0.6935   -1.14   0.25713
## Year2010          -0.0967      0.3967   -0.24   0.80791
## Year2011          -0.5153      0.3485   -1.48   0.14209
## Year2012          -0.4081      0.3402   -1.20   0.23287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.394, Adjusted R-squared:  0.303
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0931 0.8530 0.9350 0.8840 0.9880 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.63e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 3.681 1          1.919
## Year            3.681 16          1.042

```



## Residuals from last author



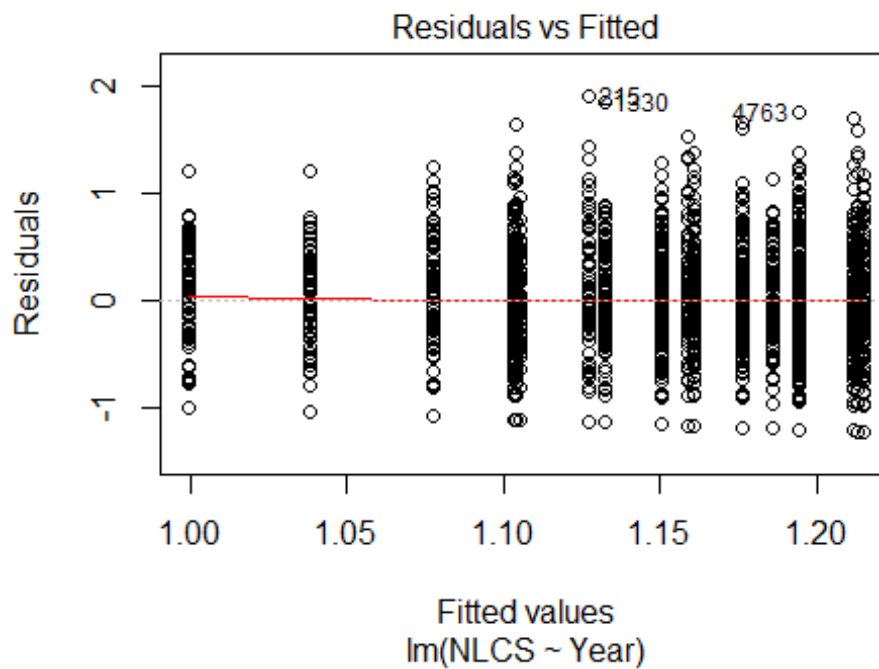
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2268 -0.3225 -0.0848 0.4702 2.3167
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.442 0.275 5.24 7.4e-07 ***
## LastAuthorFemale1 -0.879 0.153 -5.74 8.2e-08 ***
## Year1997 -0.802 0.352 -2.28 0.0246 *
## Year1998 -0.299 0.418 -0.72 0.4749
## Year1999 -0.971 0.315 -3.08 0.0026 **
## Year2000 -0.345 0.646 -0.53 0.5945
## Year2001 -0.978 0.385 -2.54 0.0124 *
## Year2002 -0.256 0.342 -0.75 0.4550
## Year2003 -0.481 0.316 -1.52 0.1309
## Year2004 -0.589 0.311 -1.90 0.0605 .
## Year2005 0.176 0.341 0.52 0.6062
## Year2006 -0.370 0.365 -1.01 0.3129
```

```

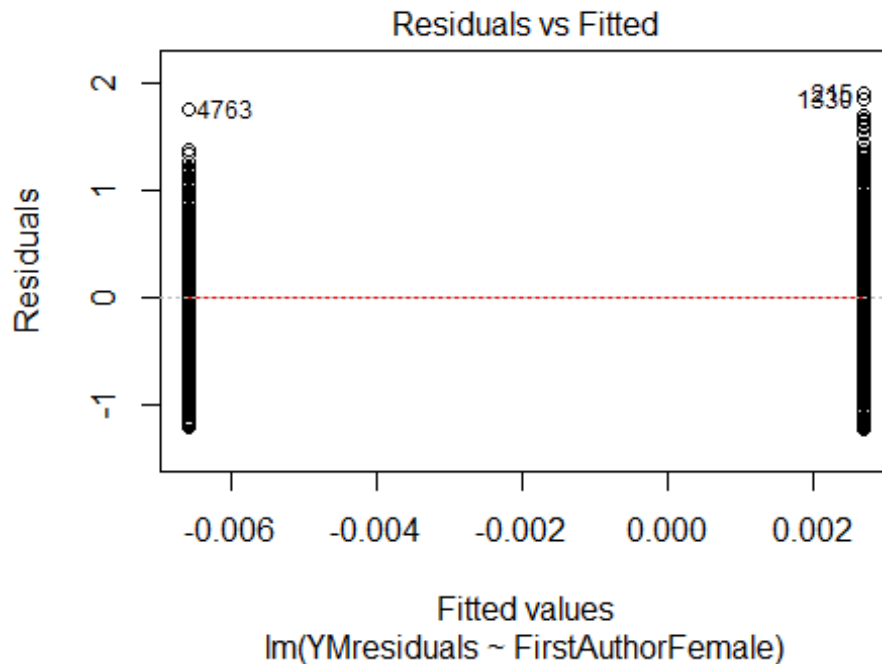
## Year2007          -0.432          0.318      -1.36      0.1768
## Year2008          -0.216          0.331      -0.65      0.5166
## Year2009          -0.352          0.329      -1.07      0.2875
## Year2010          -0.383          0.708      -0.54      0.5895
## Year2011          -0.525          0.309      -1.70      0.0917 .
## Year2012          -0.285          0.385      -0.74      0.4607
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.422, Adjusted R-squared:  0.335
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 125 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0375 0.8700 0.9370 0.8960 0.9900 0.9980
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.63e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 131"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1502"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 193 182 197 172 206 232 208 170 197 249 233 283 313 314 372
## 2011 2012
## 398 369
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 80 88 76 72 88 85 100 67 106 120 141 149 164 186 196
## 2011 2012

```

```
## 251 225
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 77 82 67 67 81 79 87 59 92 107 129 127 137 160 169
## 2011 2012
## 223 191
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 47, df = 16, p-value = 6e-05
```

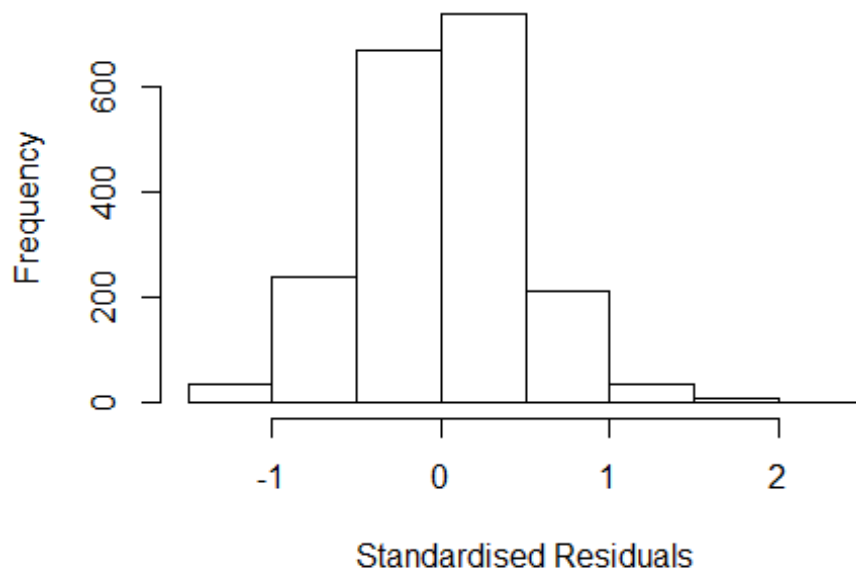


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.88097636534109"
## [1] "Male first author team size 2018 geometric mean: 3.89266811521134"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4000, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.82542039804429"
## [1] "Male last author team size 2018 geometric mean: 3.90363811463132"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.063 1          1.031
## LastAuthorFemale  1.074 1          1.036
## UniqueAuthors    1.275 4          1.031
## Year              1.334 16         1.009
```

## Residuals from first and last author and team size



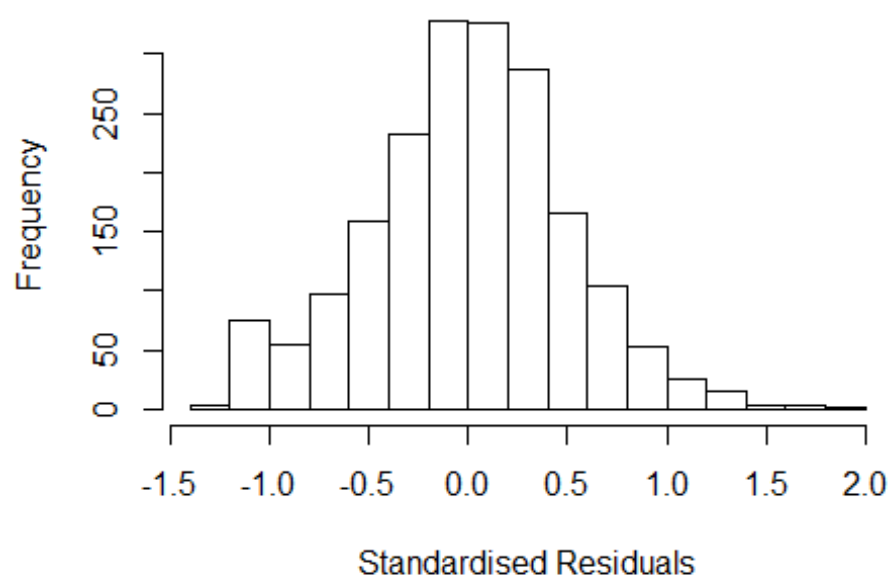
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3714 -0.2935 0.0125 0.2933 2.0674
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8019 0.0760 10.56 <2e-16 ***
## FirstAuthorFemale1 -0.0293 0.0236 -1.24 0.21
## LastAuthorFemale1 0.0128 0.0292 0.44 0.66
## UniqueAuthors2 0.3861 0.0458 8.43 <2e-16 ***
## UniqueAuthors3 0.4235 0.0417 10.15 <2e-16 ***
## UniqueAuthors4 0.4796 0.0439 10.92 <2e-16 ***
## UniqueAuthors5 0.5802 0.0403 14.39 <2e-16 ***
## Year1997 -0.0107 0.0951 -0.11 0.91
## Year1998 0.0366 0.1026 0.36 0.72
## Year1999 -0.0747 0.1000 -0.75 0.46
```

```

## Year2000          -0.1219      0.0935    -1.30      0.19
## Year2001          -0.0338      0.0926    -0.37      0.71
## Year2002          -0.1093      0.0930    -1.18      0.24
## Year2003          -0.1308      0.0903    -1.45      0.15
## Year2004           0.0200      0.0866     0.23      0.82
## Year2005          -0.0169      0.0858    -0.20      0.84
## Year2006          -0.0985      0.0820    -1.20      0.23
## Year2007          -0.0699      0.0840    -0.83      0.41
## Year2008          -0.0651      0.0832    -0.78      0.43
## Year2009          -0.0545      0.0817    -0.67      0.50
## Year2010          -0.0922      0.0807    -1.14      0.25
## Year2011          -0.0905      0.0800    -1.13      0.26
## Year2012          -0.0943      0.0813    -1.16      0.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.142, Adjusted R-squared:  0.132
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 171 weights are ~= 1. The remaining 1763 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8600 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.17e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1          1.016
## LastAuthorFemale 1.041 1          1.020
## Year              1.059 16          1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2478 -0.3134  0.0165  0.3067  1.9661
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06991    0.08569   12.49  <2e-16 ***
## FirstAuthorFemale1  0.00821    0.02458    0.33   0.738
## LastAuthorFemale1 -0.00163    0.03145   -0.05   0.959
## Year1997         0.02114    0.10989    0.19   0.847
## Year1998         0.06543    0.11534    0.57   0.571
## Year1999        -0.02425    0.10638   -0.23   0.820
## Year2000        -0.05397    0.10600   -0.51   0.611
## Year2001         0.06145    0.10958    0.56   0.575
## Year2002         0.02067    0.10768    0.19   0.848
## Year2003         0.04744    0.10113    0.47   0.639
## Year2004         0.17786    0.09654    1.84   0.066 .
## Year2005         0.16198    0.09594    1.69   0.091 .
```

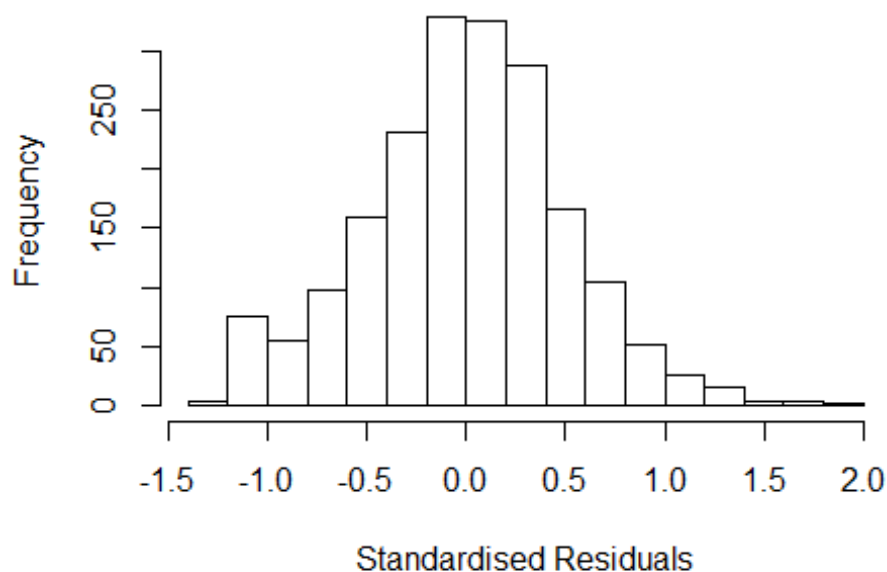
```

## Year2006          0.04126    0.09410    0.44    0.661
## Year2007          0.12800    0.09567    1.34    0.181
## Year2008          0.08279    0.09575    0.86    0.387
## Year2009          0.12169    0.09135    1.33    0.183
## Year2010          0.08349    0.09367    0.89    0.373
## Year2011          0.09857    0.09140    1.08    0.281
## Year2012          0.08814    0.09237    0.95    0.340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00315
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1776 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0391 0.8600 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1          1.012
## Year              1.024 16          1.001

```



## Residuals from first author



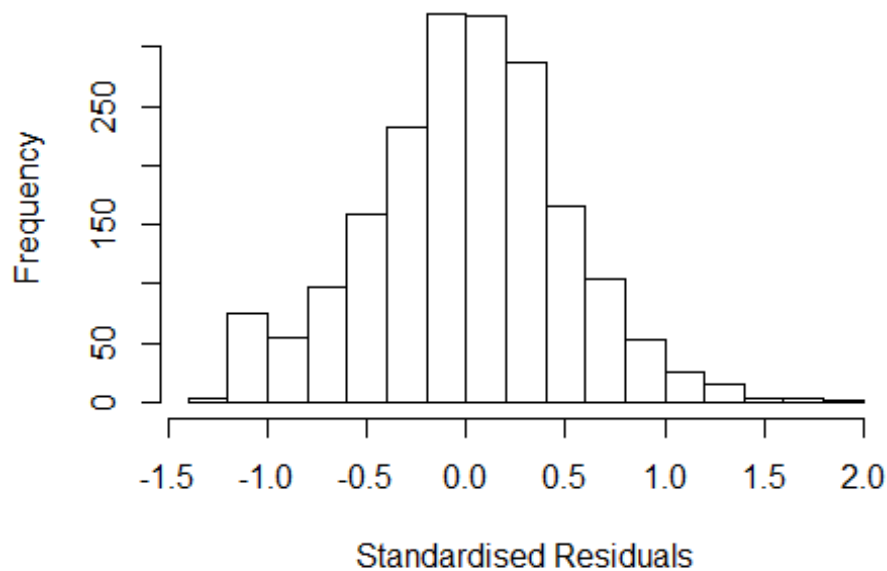
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2477 -0.3132 0.0159 0.3068 1.9662
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06984 0.08541 12.53 <2e-16 ***
## FirstAuthorFemale1 0.00798 0.02457 0.32 0.745
## Year1997 0.02114 0.10970 0.19 0.847
## Year1998 0.06533 0.11528 0.57 0.571
## Year1999 -0.02449 0.10631 -0.23 0.818
## Year2000 -0.05426 0.10595 -0.51 0.609
## Year2001 0.06118 0.10955 0.56 0.577
## Year2002 0.02050 0.10763 0.19 0.849
## Year2003 0.04729 0.10106 0.47 0.640
## Year2004 0.17781 0.09636 1.85 0.065 .
## Year2005 0.16184 0.09589 1.69 0.092 .
## Year2006 0.04108 0.09404 0.44 0.662
```

```

## Year2007          0.12774      0.09561      1.34      0.182
## Year2008          0.08257      0.09569      0.86      0.388
## Year2009          0.12160      0.09121      1.33      0.183
## Year2010          0.08330      0.09359      0.89      0.374
## Year2011          0.09842      0.09134      1.08      0.281
## Year2012          0.08807      0.09230      0.95      0.340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00366
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 156 weights are ~= 1. The remaining 1778 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0404 0.8610 0.9520 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1      1.017
## Year              1.035 16      1.001

```

## Residuals from last author



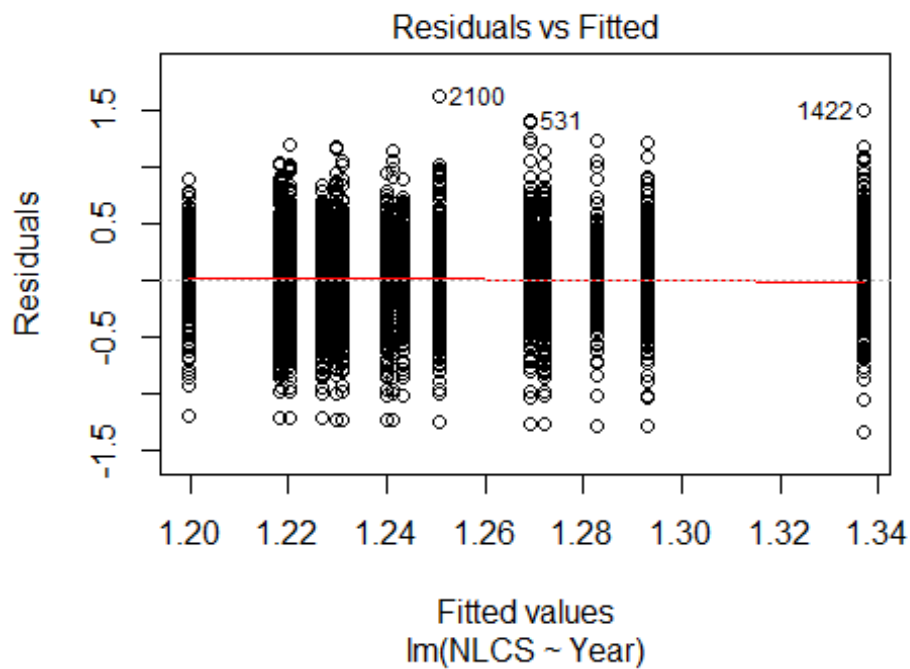
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2504 -0.3145  0.0181  0.3079  1.9644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07e+00  8.55e-02  12.53  <2e-16 ***
## LastAuthorFemale1 -1.71e-05  3.14e-02   0.00   1.000
## Year1997         2.16e-02  1.10e-01   0.20   0.844
## Year1998         6.55e-02  1.15e-01   0.57   0.570
## Year1999        -2.38e-02  1.06e-01  -0.22   0.823
## Year2000        -5.36e-02  1.06e-01  -0.50   0.614
## Year2001         6.24e-02  1.10e-01   0.57   0.569
## Year2002         2.14e-02  1.08e-01   0.20   0.842
## Year2003         4.86e-02  1.01e-01   0.48   0.631
## Year2004         1.79e-01  9.66e-02   1.85   0.064 .
## Year2005         1.63e-01  9.59e-02   1.70   0.090 .
## Year2006         4.17e-02  9.42e-02   0.44   0.658
```

```

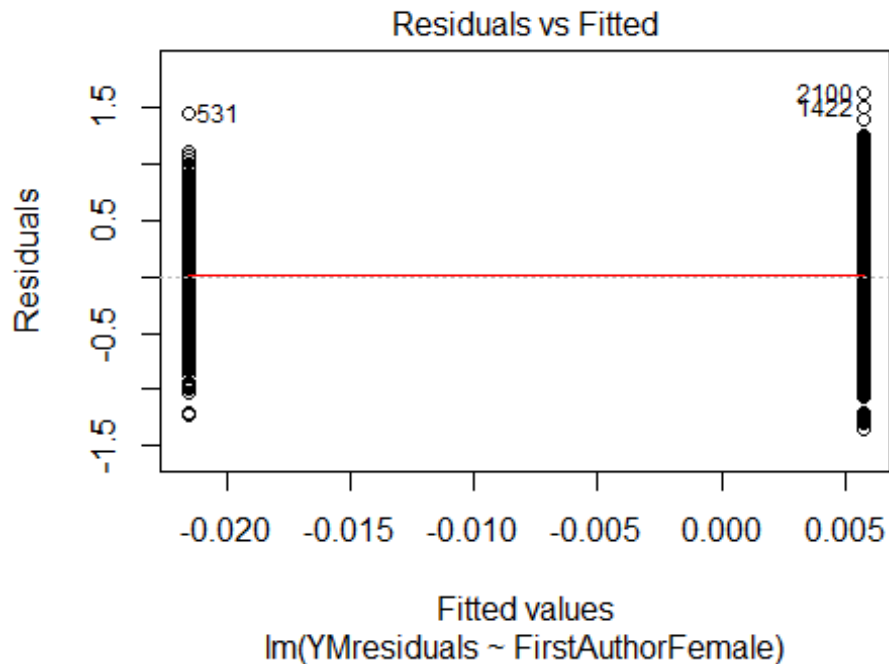
## Year2007      1.28e-01  9.57e-02  1.34  0.181
## Year2008      8.28e-02  9.58e-02  0.86  0.387
## Year2009      1.22e-01  9.13e-02  1.34  0.180
## Year2010      8.40e-02  9.37e-02  0.90  0.370
## Year2011      9.91e-02  9.14e-02  1.08  0.278
## Year2012      8.84e-02  9.24e-02  0.96  0.339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00364
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 152 weights are ~= 1. The remaining 1782 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0392 0.8610 0.9520 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1934"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1503"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 476 410 441 405 459 426 464 438 439 426 426 413 413 489 447
## 2011 2012
## 530 477
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 344 311 337 302 243 173 355 325 329 339 317 314 311 364 342
## 2011 2012

```

```
## 403 372
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 292 263 299 275 218 153 322 294 305 297 294 281 290 331 302
## 2011 2012
## 365 341
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 65, df = 16, p-value = 6e-08
```

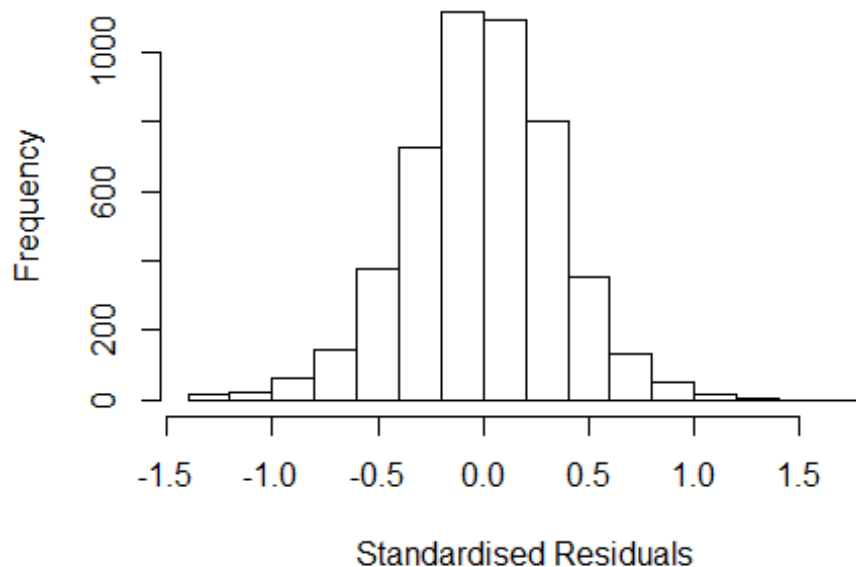


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.8, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 5.08065332861046"
## [1] "Male first author team size 2018 geometric mean: 4.51846290709934"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.97801996388118"
## [1] "Male last author team size 2018 geometric mean: 4.62137780079086"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## LastAuthorFemale  1.021 1          1.010
## UniqueAuthors    1.110 4          1.013
## Year             1.146 16          1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.344159 -0.222639 -0.000654 0.226607 1.642844
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18221 0.04204 28.12 <2e-16 ***
## FirstAuthorFemale1 -0.03291 0.01215 -2.71 0.0068 **
## LastAuthorFemale1 -0.01224 0.01579 -0.78 0.4380
## UniqueAuthors2 0.06105 0.04087 1.49 0.1353
## UniqueAuthors3 0.05997 0.03970 1.51 0.1310
## UniqueAuthors4 0.12670 0.03968 3.19 0.0014 **
## UniqueAuthors5 0.12440 0.03898 3.19 0.0014 **
## Year1997 0.00949 0.03154 0.30 0.7637
## Year1998 0.05324 0.03192 1.67 0.0954 .
## Year1999 -0.01411 0.03151 -0.45 0.6542
```

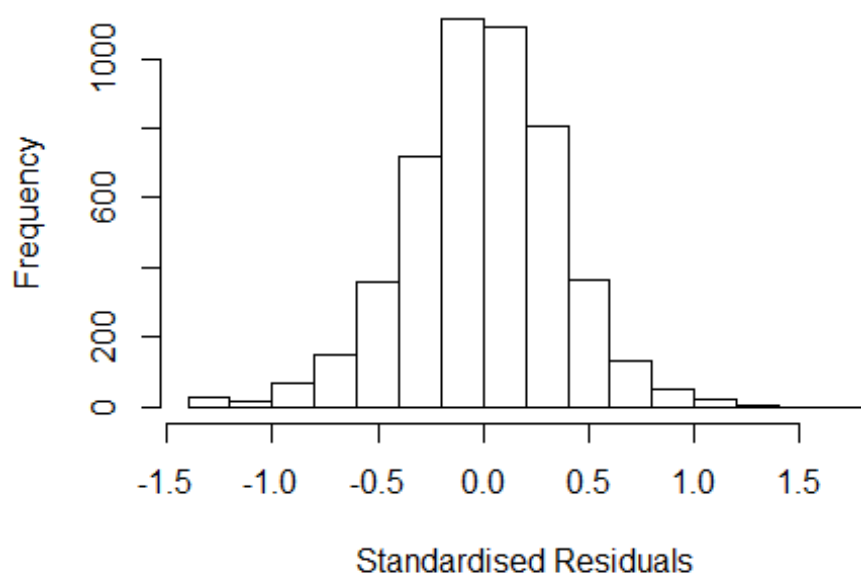
```

## Year2000      -0.00515    0.03277   -0.16    0.8751
## Year2001      0.03755    0.03432    1.09    0.2740
## Year2002     -0.03081    0.03054   -1.01    0.3131
## Year2003     -0.05267    0.02853   -1.85    0.0649 .
## Year2004     -0.04971    0.02893   -1.72    0.0858 .
## Year2005     -0.05523    0.02998   -1.84    0.0655 .
## Year2006     -0.02715    0.02807   -0.97    0.3335
## Year2007     -0.04077    0.02831   -1.44    0.1500
## Year2008     -0.02440    0.02845   -0.86    0.3912
## Year2009     -0.05443    0.02850   -1.91    0.0562 .
## Year2010     -0.04045    0.02897   -1.40    0.1627
## Year2011      0.01348    0.02810    0.48    0.6314
## Year2012     -0.02162    0.02810   -0.77    0.4417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.338
## Multiple R-squared:  0.0186, Adjusted R-squared:  0.0142
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1127 is an outlier with |weight| = 0 ( < 2e-05);
## 403 weights are ~= 1. The remaining 4518 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0174 0.8690 0.9520 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.03e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1 1.016
## LastAuthorFemale 1.015 1 1.007
## Year 1.044 16 1.001

```



## Residuals from first and last author



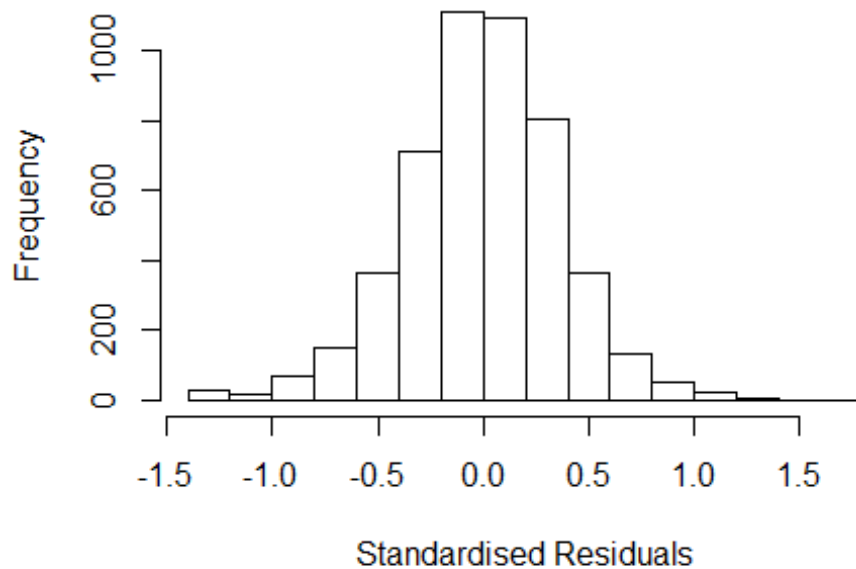
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32546 -0.22550  0.00109  0.22745  1.61346
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.271087   0.020927  60.74  <2e-16 ***
## FirstAuthorFemale1 -0.030005   0.012196  -2.46   0.014 *
## LastAuthorFemale1 -0.009089   0.015873  -0.57   0.567
## Year1997         0.013750   0.031588   0.44   0.663
## Year1998         0.054370   0.031857   1.71   0.088 .
## Year1999        -0.012550   0.031572  -0.40   0.691
## Year2000        -0.000394   0.032730  -0.01   0.990
## Year2001         0.038942   0.034526   1.13   0.259
## Year2002        -0.029588   0.030388  -0.97   0.330
## Year2003        -0.046414   0.028340  -1.64   0.102
## Year2004        -0.047593   0.028818  -1.65   0.099 .
## Year2005        -0.046634   0.030035  -1.55   0.121
```

```

## Year2006      -0.014751    0.028063   -0.53    0.599
## Year2007      -0.029765    0.028380   -1.05    0.294
## Year2008      -0.011683    0.028296   -0.41    0.680
## Year2009      -0.045402    0.028614   -1.59    0.113
## Year2010      -0.027895    0.028796   -0.97    0.333
## Year2011       0.023374    0.027946    0.84    0.403
## Year2012      -0.008445    0.028069   -0.30    0.764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.00837,    Adjusted R-squared:  0.00473
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1127 is an outlier with |weight| = 0 ( < 2e-05);
## 406 weights are ~= 1. The remaining 4515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0114 0.8710 0.9510 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.03 1          1.015
## Year              1.03 16          1.001

```

## Residuals from first author



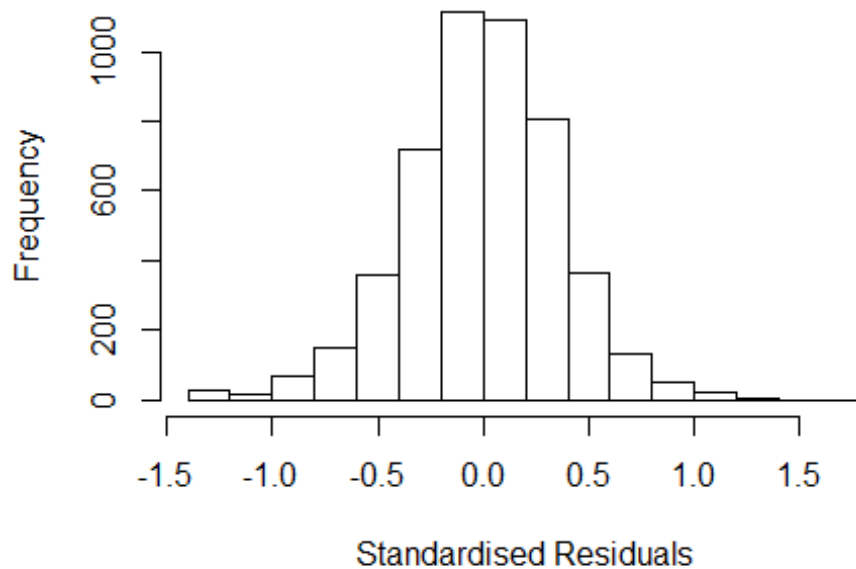
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32434 -0.22640 0.00136 0.22781 1.61437
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.270327 0.020878 60.84 <2e-16 ***
## FirstAuthorFemale1 -0.030038 0.012197 -2.46 0.014 *
## Year1997 0.013774 0.031586 0.44 0.663
## Year1998 0.054017 0.031852 1.70 0.090 .
## Year1999 -0.012694 0.031558 -0.40 0.688
## Year2000 -0.000387 0.032733 -0.01 0.991
## Year2001 0.038409 0.034505 1.11 0.266
## Year2002 -0.030016 0.030363 -0.99 0.323
## Year2003 -0.046815 0.028353 -1.65 0.099 .
## Year2004 -0.047823 0.028812 -1.66 0.097 .
## Year2005 -0.046935 0.030048 -1.56 0.118
## Year2006 -0.015229 0.028056 -0.54 0.587
```

```

## Year2007          -0.029755    0.028376   -1.05    0.294
## Year2008          -0.012047    0.028302   -0.43    0.670
## Year2009          -0.045973    0.028618   -1.61    0.108
## Year2010          -0.028185    0.028798   -0.98    0.328
## Year2011           0.023063    0.027952    0.83    0.409
## Year2012          -0.008515    0.028078   -0.30    0.762
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.00829,    Adjusted R-squared:  0.00485
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1127 is an outlier with |weight| = 0 ( < 2e-05);
## 414 weights are ~= 1. The remaining 4507 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0111 0.8710 0.9510 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.013 1          1.006
## Year            1.013 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32049 -0.22548 0.00148 0.23077 1.61746
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26793 0.02093 60.58 <2e-16 ***
## LastAuthorFemale1 -0.00928 0.01587 -0.58 0.559
## Year1997 0.01198 0.03157 0.38 0.704
## Year1998 0.05256 0.03183 1.65 0.099 .
## Year1999 -0.01340 0.03162 -0.42 0.672
## Year2000 -0.00141 0.03269 -0.04 0.966
## Year2001 0.03592 0.03456 1.04 0.299
## Year2002 -0.03293 0.03041 -1.08 0.279
## Year2003 -0.05014 0.02831 -1.77 0.077 .
## Year2004 -0.05022 0.02885 -1.74 0.082 .
## Year2005 -0.04993 0.03003 -1.66 0.096 .
## Year2006 -0.01879 0.02801 -0.67 0.502
```

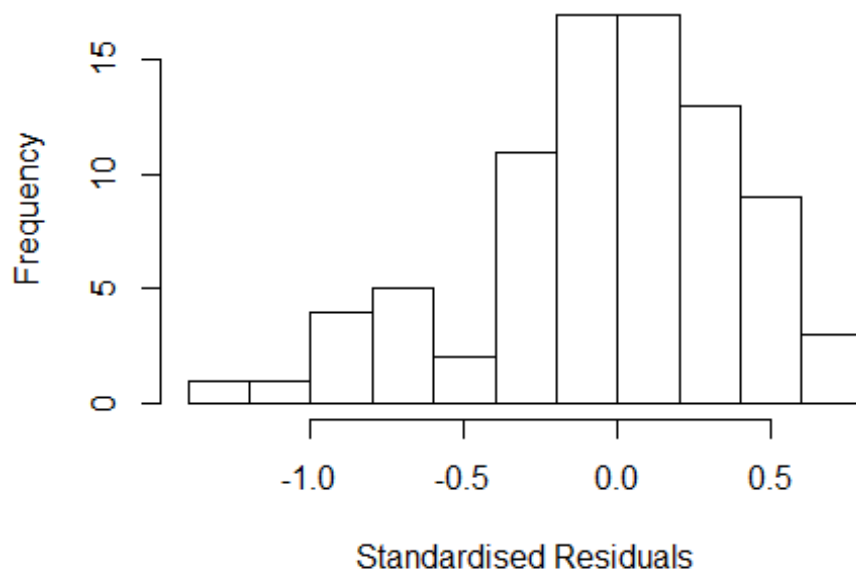
```

## Year2007          -0.03319      0.02834    -1.17      0.242
## Year2008          -0.01745      0.02816    -0.62      0.536
## Year2009          -0.04969      0.02859    -1.74      0.082 .
## Year2010          -0.03122      0.02881    -1.08      0.279
## Year2011           0.01655      0.02781      0.59      0.552
## Year2012          -0.01345      0.02803    -0.48      0.631
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.00711,    Adjusted R-squared:  0.00367
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1127 is an outlier with |weight| = 0 ( < 2e-05);
## 428 weights are ~= 1. The remaining 4493 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0104 0.8700 0.9510 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4922"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1504"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10   8  12  11  17  12  12   9  17  12  16  14  15   3  10
## 2011 2012
##    3   5
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    3   4   3   6   8   5   5   6  10   5  12   7   8   0   8

```

```
## 2011 2012
##    1    3
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    4    3    5    7    5    4    4    6    5   12    6    8    0    8
## 2011 2012
##    1    3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: 3.46410161513775"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale    7.732  1         2.781
## LastAuthorFemale   13.469  1         3.670
## UniqueAuthors    1469.950  4         2.488
## Year              7043.755 15         1.344
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
```

```

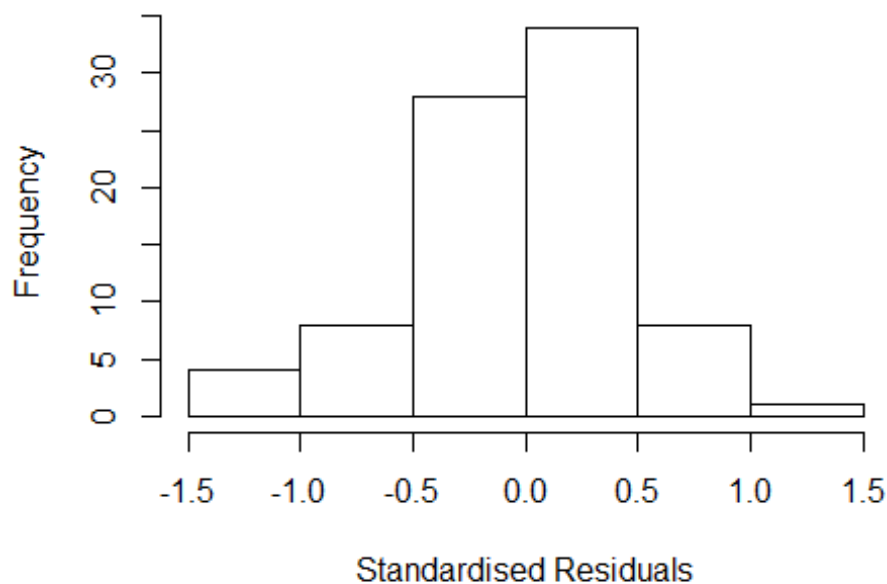
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.2040 -0.2671  0.0398  0.2859  0.6801
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0442     0.1921   5.43   1e-06 ***
## FirstAuthorFemale1  0.0698     0.1384   0.50  0.61582
## LastAuthorFemale1 -0.4122     0.1758  -2.34  0.02232 *
## UniqueAuthors2     0.4950     0.1841   2.69  0.00924 **
## UniqueAuthors3     0.2163     0.2020   1.07  0.28849
## UniqueAuthors4    -0.0627     0.2766  -0.23  0.82129
## UniqueAuthors5     0.2534     0.3043   0.83  0.40823
## Year1997          -0.5895     0.2579  -2.29  0.02572 *
## Year1998          -0.1613     0.1876  -0.86  0.39312
## Year1999          -0.3395     0.2517  -1.35  0.18239
## Year2000          -0.2133     0.2854  -0.75  0.45771
## Year2001           0.3468     0.2150   1.61  0.11192
## Year2002          -0.3581     0.3164  -1.13  0.26222
## Year2003           0.3458     0.3038   1.14  0.25947
## Year2004           0.3824     0.2745   1.39  0.16857
## Year2005          -0.0870     0.3920  -0.22  0.82507
## Year2006          -0.3191     0.1405  -2.27  0.02671 *
## Year2007          -0.2213     0.2073  -1.07  0.29010
## Year2008          -0.1944     0.1805  -1.08  0.28577
## Year2010           0.1334     0.2436   0.55  0.58599
## Year2011           0.9815     0.2624   3.74  0.00041 ***
## Year2012          -0.3244     0.2595  -1.25  0.21610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.392, Adjusted R-squared:  0.182
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~ = 1. The remaining 77 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.441  0.876  0.958  0.913  0.992  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale

```



```
##           500           50           2           1           1000           200
## trace.lev      mts compute.rd
##           0           1000           0
##           psi      subsampling      cov
##           "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.68 1 2.585
## LastAuthorFemale 10.44 1 3.232
## Year 17.70 15 1.101
```

### Residuals from first and last author



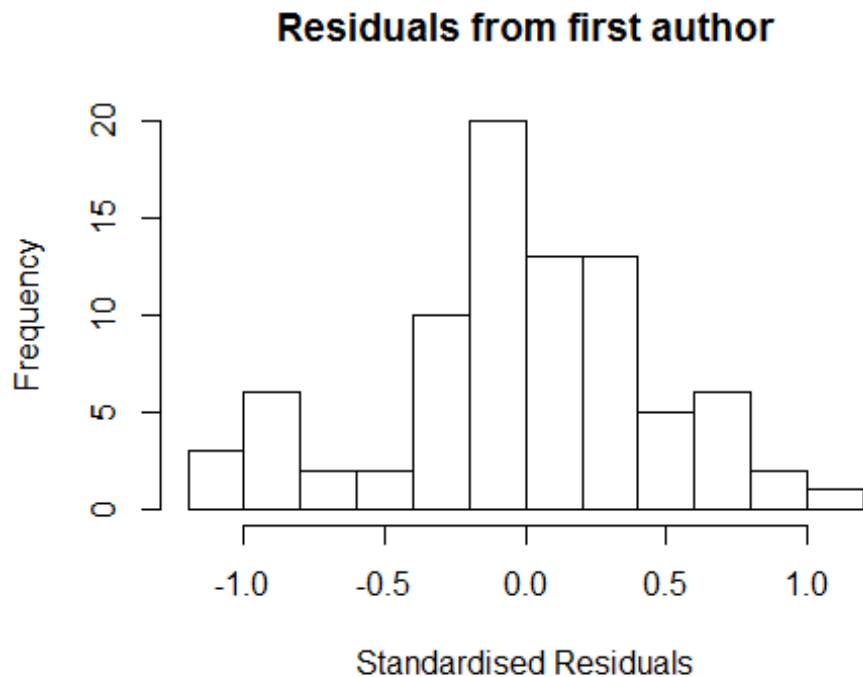
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3487 -0.2109 0.0146 0.2581 1.0290
```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.5104    0.0883   17.10 < 2e-16 ***
## FirstAuthorFemale1  0.1056    0.1214    0.87  0.3876
## LastAuthorFemale1 -0.3904    0.1572   -2.48  0.0156 *
## Year1997        -0.5891    0.3327   -1.77  0.0813 .
## Year1998        -0.1571    0.1886   -0.83  0.4080
## Year1999        -0.6052    0.1984   -3.05  0.0033 **
## Year2000        -0.4516    0.2737   -1.65  0.1038
## Year2001         0.0645    0.2102    0.31  0.7601
## Year2002        -0.7794    0.3922   -1.99  0.0511 .
## Year2003         0.0182    0.2136    0.09  0.9323
## Year2004        -0.0389    0.2348   -0.17  0.8690
## Year2005        -0.4437    0.3571   -1.24  0.2184
## Year2006        -0.5616    0.1108   -5.07  3.6e-06 ***
## Year2007        -0.4167    0.2098   -1.99  0.0513 .
## Year2008        -0.4035    0.1754   -2.30  0.0247 *
## Year2010        -0.2673    0.1770   -1.51  0.1359
## Year2011         0.4526    0.0883    5.13  2.9e-06 ***
## Year2012        -0.5846    0.2615   -2.24  0.0288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.277, Adjusted R-squared:  0.0878
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 74 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.361  0.883  0.967  0.900  0.990  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))

```

```
## FirstAuthorFemale 2.908 1 1.705
## Year 2.908 15 1.036
```



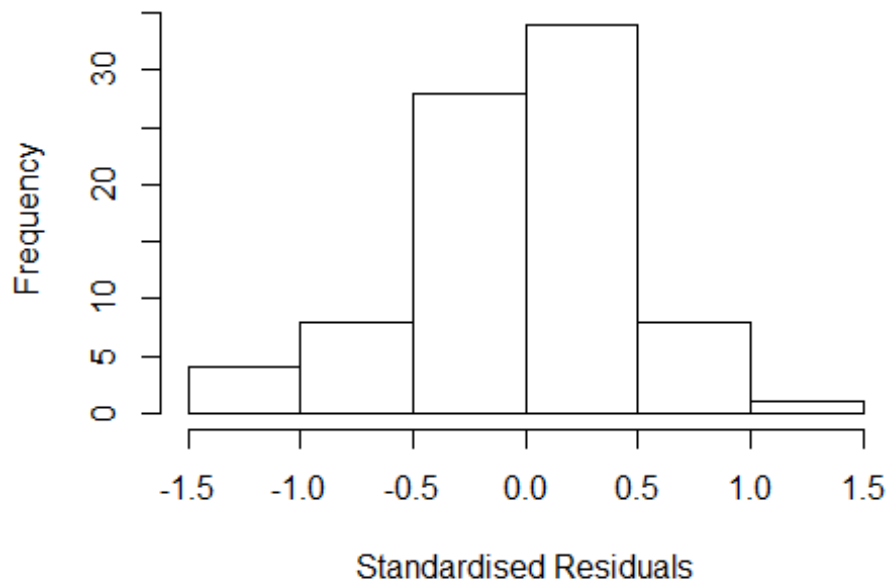
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0774 -0.2194 -0.0163 0.3221 1.0196
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36522 0.25338 5.39 1e-06 ***
## FirstAuthorFemale1 0.00555 0.11440 0.05 0.961
## Year1997 -0.41844 0.38294 -1.09 0.278
## Year1998 0.05710 0.27432 0.21 0.836
## Year1999 -0.51820 0.32548 -1.59 0.116
## Year2000 -0.28792 0.35049 -0.82 0.414
## Year2001 0.22867 0.30321 0.75 0.453
## Year2002 -0.62486 0.45206 -1.38 0.172
## Year2003 0.07504 0.33303 0.23 0.822
```

```

## Year2004          0.10294      0.33454      0.31      0.759
## Year2005          -0.23601      0.43909     -0.54      0.593
## Year2006          -0.48117      0.26195     -1.84      0.071 .
## Year2007          -0.24206      0.31384     -0.77      0.443
## Year2008          -0.28780      0.32217     -0.89      0.375
## Year2010          -0.34857      0.31972     -1.09      0.280
## Year2011           0.59778      0.25338       2.36      0.021 *
## Year2012          -0.55580      0.37495     -1.48      0.143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.48
## Multiple R-squared:  0.227, Adjusted R-squared:  0.0397
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 78 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.594  0.866  0.959   0.910   0.991   0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 6.44 1          2.538
## Year             6.44 15          1.064

```

## Residuals from last author



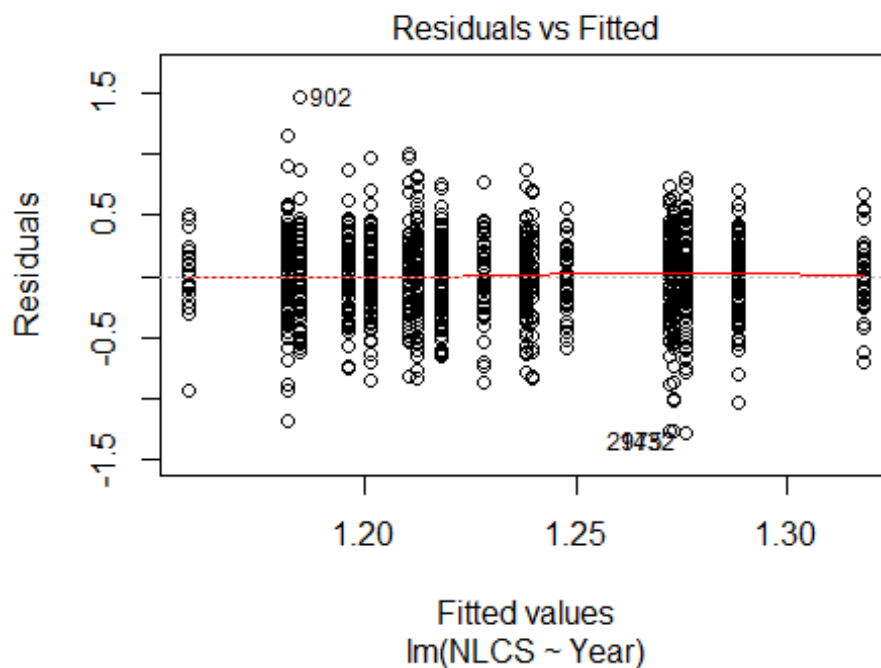
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2567 -0.2107 0.0371 0.2846 1.0315
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.53839 0.13234 11.62 < 2e-16 ***
## LastAuthorFemale1 -0.34077 0.14646 -2.33 0.02306 *
## Year1997 -0.57680 0.31334 -1.84 0.07015 .
## Year1998 -0.11251 0.17622 -0.64 0.52538
## Year1999 -0.59290 0.22731 -2.61 0.01124 *
## Year2000 -0.45380 0.28718 -1.58 0.11884
## Year2001 0.05547 0.22097 0.25 0.80258
## Year2002 -0.80987 0.40690 -1.99 0.05070 .
## Year2003 0.00791 0.23634 0.03 0.97340
## Year2004 -0.06602 0.25367 -0.26 0.79547
## Year2005 -0.40748 0.40078 -1.02 0.31300
## Year2006 -0.55294 0.14642 -3.78 0.00034 ***
```

```

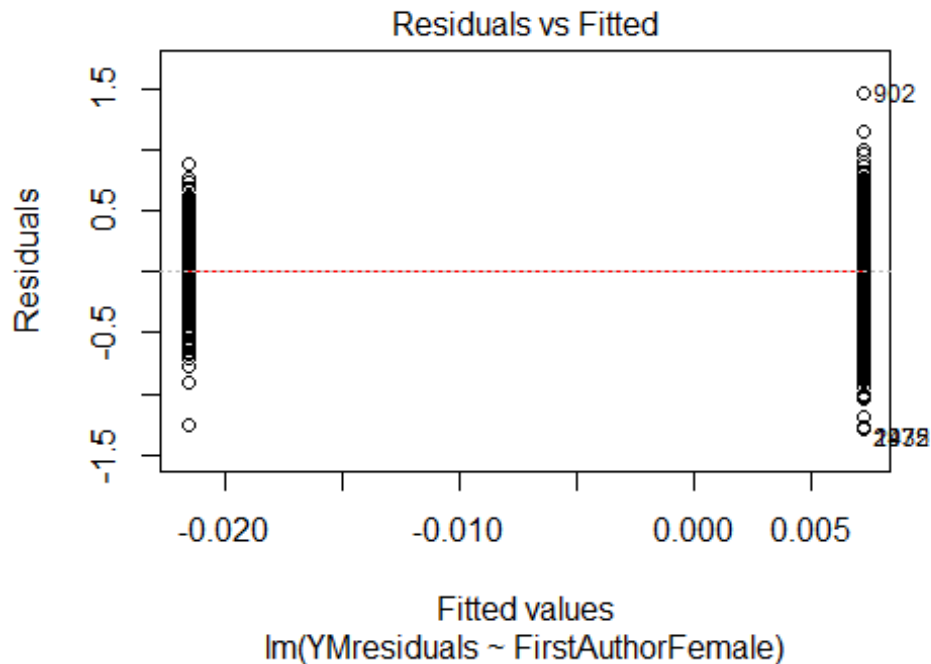
## Year2007      -0.40842    0.23029   -1.77  0.08076 .
## Year2008      -0.41944    0.20922   -2.00  0.04909 *
## Year2010      -0.28170    0.20467   -1.38  0.17337
## Year2011       0.42461    0.13234    3.21  0.00206 **
## Year2012      -0.58940    0.27749   -2.12  0.03742 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.271, Adjusted R-squared:  0.0938
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 75 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.415  0.888  0.960  0.898  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 83"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1505"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  106  83  105  109  95  86  126  144  137  137  164  125  144  155  142
## 2011 2012
##  144  141
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47  35  40  57  44  22  76  85  91  99  96  70  89  106  97
## 2011 2012
##   98  99

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   41   33   37   53   42   21   62   81   85   86   84   63   80   93   84
## 2011 2012
##   86   90
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```



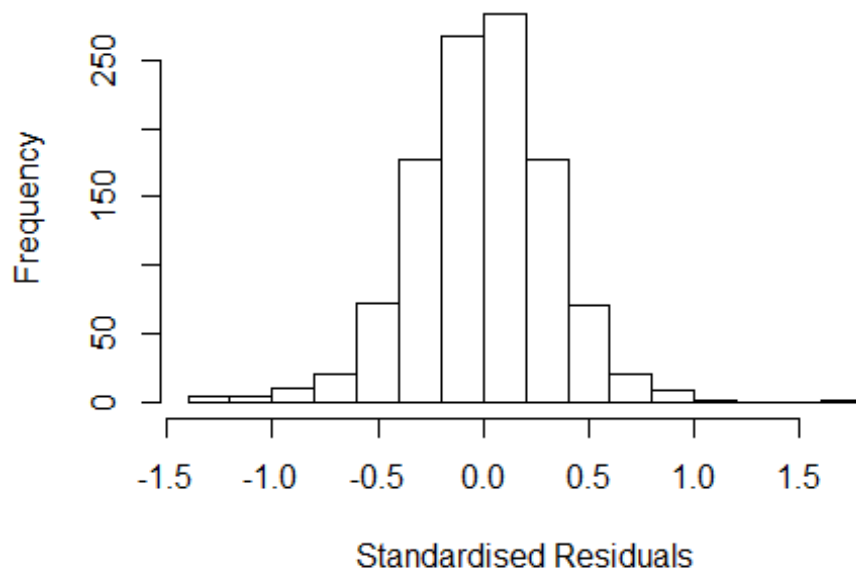
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.41, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 6.13153803909595"
## [1] "Male first author team size 2018 geometric mean: 5.34172853865677"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 620, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.71616754293713"
## [1] "Male last author team size 2018 geometric mean: 5.50052125860327"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 370, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.069 1      1.034
## LastAuthorFemale  1.071 1      1.035
## UniqueAuthors    1.459 4      1.048
## Year             1.600 16      1.015
```



## Residuals from first and last author and team size



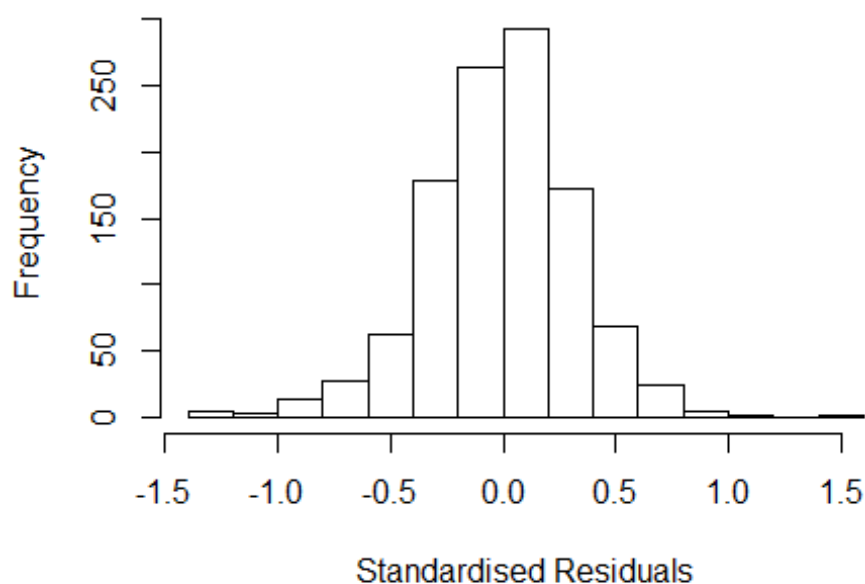
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32095 -0.20963  0.00152  0.19864  1.61441
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.15334    0.06067   19.01  <2e-16 ***
## FirstAuthorFemale1 -0.04316    0.02187   -1.97   0.0487 *
## LastAuthorFemale1 -0.00657    0.02878   -0.23   0.8194
## UniqueAuthors2     0.07204    0.05885    1.22   0.2212
## UniqueAuthors3     0.12053    0.05418    2.22   0.0263 *
## UniqueAuthors4     0.17098    0.05439    3.14   0.0017 **
## UniqueAuthors5     0.19370    0.05341    3.63   0.0003 ***
## Year1997         -0.03005    0.07774   -0.39   0.6992
## Year1998          0.05313    0.06130    0.87   0.3863
## Year1999         -0.06306    0.06218   -1.01   0.3107
```

```

## Year2000      0.00782    0.06367    0.12    0.9022
## Year2001     -0.08050    0.07053   -1.14    0.2539
## Year2002     -0.11875    0.05595   -2.12    0.0340 *
## Year2003     -0.08746    0.05657   -1.55    0.1224
## Year2004     -0.08876    0.05358   -1.66    0.0979 .
## Year2005     -0.00336    0.05446   -0.06    0.9508
## Year2006     -0.05127    0.05192   -0.99    0.3237
## Year2007     -0.10084    0.05659   -1.78    0.0750 .
## Year2008     -0.06362    0.05823   -1.09    0.2748
## Year2009     -0.07604    0.04913   -1.55    0.1220
## Year2010     -0.00796    0.05176   -0.15    0.8778
## Year2011      0.03614    0.05217    0.69    0.4886
## Year2012      0.02761    0.05416    0.51    0.6103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.301
## Multiple R-squared:  0.0608, Adjusted R-squared:  0.042
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 280 is an outlier with |weight| = 0 ( < 8.9e-05);
## 92 weights are ~= 1. The remaining 1028 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0156 0.8730 0.9530 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1      1.028
## LastAuthorFemale 1.059 1      1.029
## Year      1.119 16      1.004

```

## Residuals from first and last author



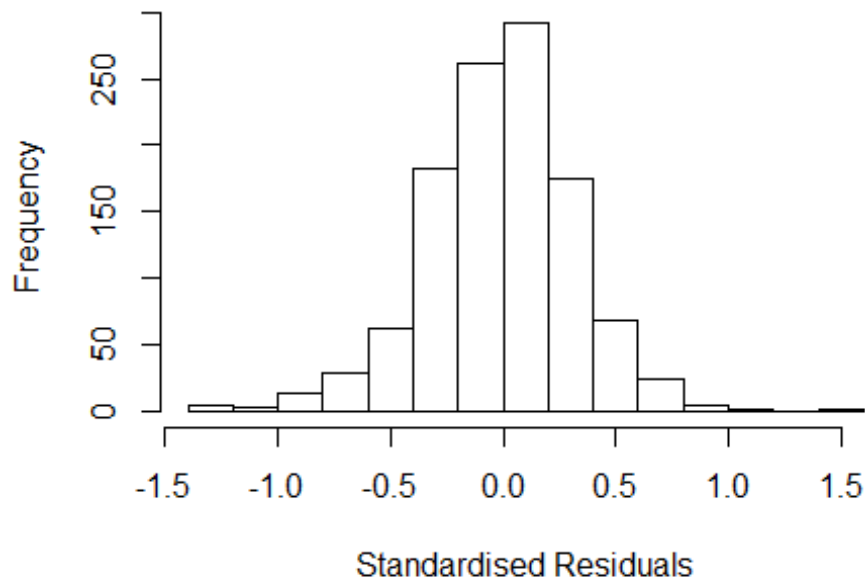
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34858 -0.21023 0.00343 0.19554 1.48024
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27091 0.04258 29.85 <2e-16 ***
## FirstAuthorFemale1 -0.03945 0.02218 -1.78 0.076 .
## LastAuthorFemale1 -0.00837 0.02856 -0.29 0.769
## Year1997 -0.00527 0.07933 -0.07 0.947
## Year1998 0.05147 0.06142 0.84 0.402
## Year1999 -0.07028 0.06445 -1.09 0.276
## Year2000 0.00710 0.06477 0.11 0.913
## Year2001 -0.07969 0.07191 -1.11 0.268
## Year2002 -0.10215 0.05864 -1.74 0.082 .
## Year2003 -0.06694 0.05643 -1.19 0.236
## Year2004 -0.07104 0.05527 -1.29 0.199
## Year2005 0.02001 0.05661 0.35 0.724
```

```

## Year2006      -0.01404    0.05346   -0.26    0.793
## Year2007      -0.06044    0.05797   -1.04    0.297
## Year2008      -0.02542    0.05951   -0.43    0.669
## Year2009      -0.03134    0.04993   -0.63    0.530
## Year2010       0.03232    0.05277    0.61    0.540
## Year2011       0.07258    0.05323    1.36    0.173
## Year2012       0.07767    0.05444    1.43    0.154
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.304
## Multiple R-squared:  0.0329, Adjusted R-squared:  0.0171
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 280 is an outlier with |weight| = 0 ( < 8.9e-05);
## 94 weights are ~= 1. The remaining 1026 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0113 0.8720 0.9520 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.057 1          1.028
## Year              1.057 16          1.002

```

## Residuals from first author



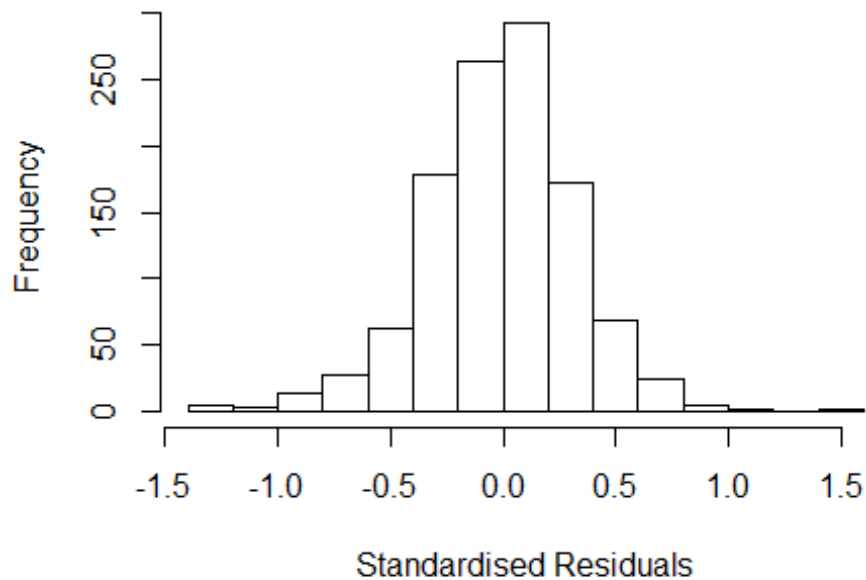
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34798 -0.20922 0.00414 0.19632 1.48149
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26975 0.04228 30.03 <2e-16 ***
## FirstAuthorFemale1 -0.03953 0.02219 -1.78 0.075 .
## Year1997 -0.00479 0.07945 -0.06 0.952
## Year1998 0.05141 0.06155 0.84 0.404
## Year1999 -0.07051 0.06445 -1.09 0.274
## Year2000 0.00681 0.06478 0.11 0.916
## Year2001 -0.08092 0.07196 -1.12 0.261
## Year2002 -0.10224 0.05861 -1.74 0.081 .
## Year2003 -0.06697 0.05651 -1.19 0.236
## Year2004 -0.07094 0.05527 -1.28 0.200
## Year2005 0.02051 0.05652 0.36 0.717
## Year2006 -0.01428 0.05354 -0.27 0.790
```

```

## Year2007          -0.06014      0.05794      -1.04      0.300
## Year2008          -0.02546      0.05956      -0.43      0.669
## Year2009          -0.03108      0.04993      -0.62      0.534
## Year2010           0.03246      0.05276       0.62      0.539
## Year2011           0.07311      0.05315       1.38      0.169
## Year2012           0.07823      0.05430       1.44      0.150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.305
## Multiple R-squared:  0.0328, Adjusted R-squared:  0.0179
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 280 is an outlier with |weight| = 0 ( < 8.9e-05);
## 91 weights are ~= 1. The remaining 1029 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0118 0.8720 0.9520 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.059 1          1.029
## Year              1.059 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3365 -0.2078  0.0044  0.1990  1.4870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.26393    0.04169   30.32  <2e-16 ***
## LastAuthorFemale1 -0.00899    0.02867   -0.31    0.754
## Year1997        -0.01077    0.07991   -0.13    0.893
## Year1998         0.05124    0.06106    0.84    0.402
## Year1999        -0.07239    0.06378   -1.13    0.257
## Year2000         0.00837    0.06401    0.13    0.896
## Year2001        -0.08635    0.07104   -1.22    0.224
## Year2002        -0.10196    0.05878   -1.73    0.083 .
## Year2003        -0.07108    0.05606   -1.27    0.205
## Year2004        -0.07332    0.05493   -1.33    0.182
## Year2005         0.01926    0.05620    0.34    0.732
## Year2006        -0.01895    0.05287   -0.36    0.720
```

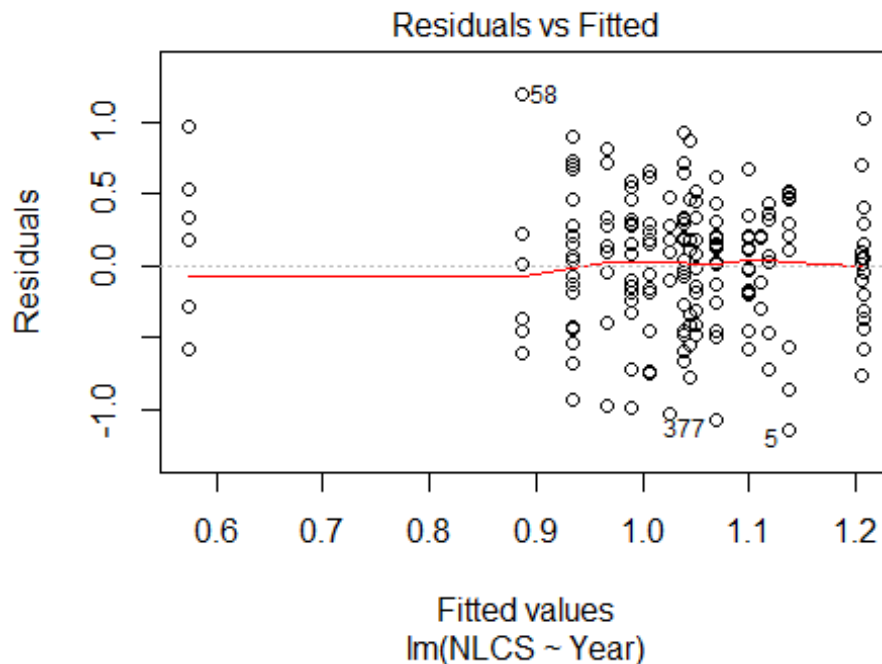
```

## Year2007      -0.06290      0.05737      -1.10      0.273
## Year2008      -0.03125      0.05866      -0.53      0.594
## Year2009      -0.03587      0.04929      -0.73      0.467
## Year2010       0.02970      0.05231       0.57      0.570
## Year2011       0.06926      0.05288       1.31      0.191
## Year2012       0.07262      0.05382       1.35      0.178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.304
## Multiple R-squared:  0.0299, Adjusted R-squared:  0.015
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 280 is an outlier with |weight| = 0 ( < 8.9e-05);
## 91 weights are ~= 1. The remaining 1029 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0147 0.8700 0.9520 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1121"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1506"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32   33   24   23   29   38   42   18   16   25   49   20   29   34   21
## 2011 2012
##   27   23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10    6    9    7    8   13   21   10    4   11   16    7   15   21   11

```



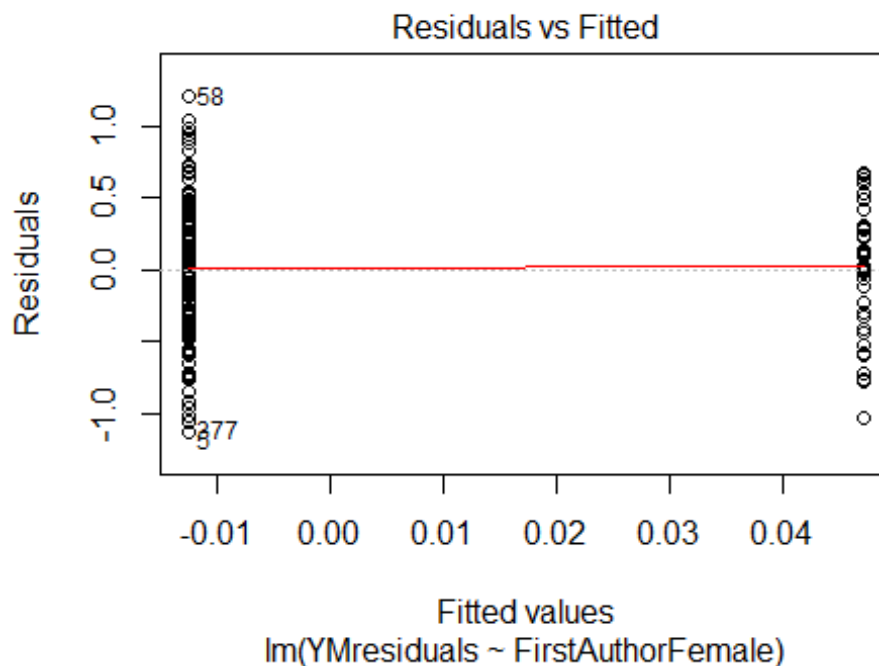
```
## 2011 2012
## 15 6
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 6 9 5 7 12 16 9 3 6 12 7 13 18 7
## 2011 2012
## 13 5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.003, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 6.7694724277124"
## [1] "Male first author team size 2018 geometric mean: 3.65905165331721"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.95789213552895"
## [1] "Male last author team size 2018 geometric mean: 3.98422018965845"

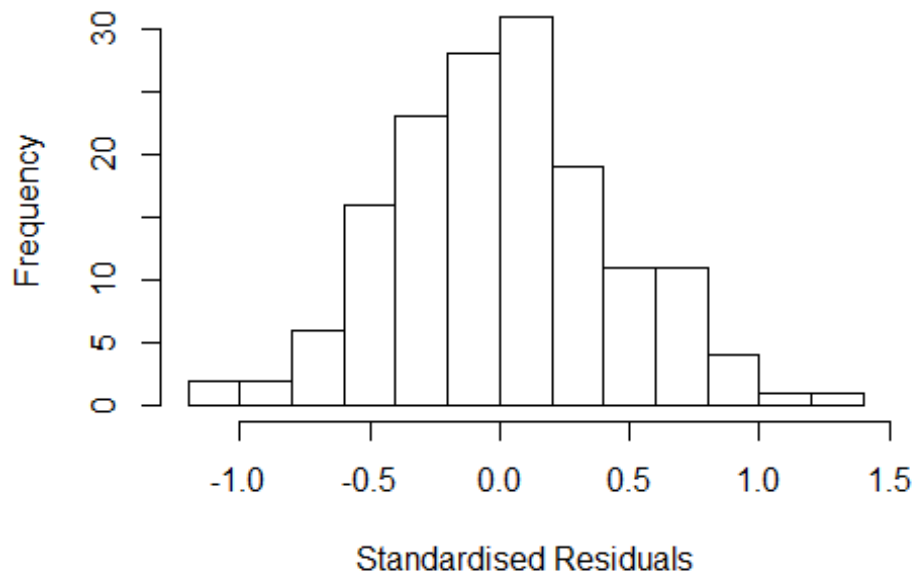
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
## FirstAuthorFemale	1.760	1	1.326
## LastAuthorFemale	2.515	1	1.586
## UniqueAuthors	11.672	4	1.360
## Year	29.621	16	1.112

## Residuals from first and last author and team size



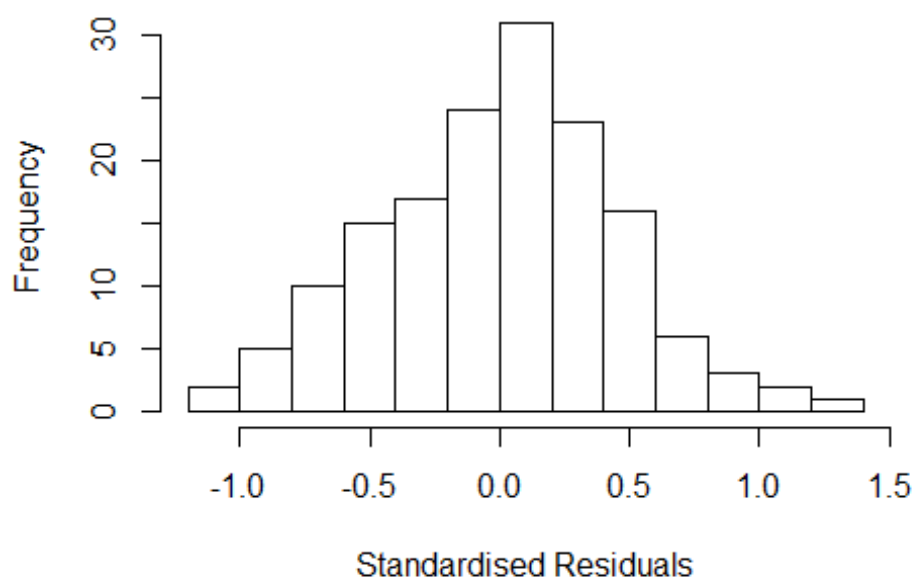
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0821 -0.2696 0.0051 0.2804 1.3746
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6963 0.3104 2.24 0.02654 *
## FirstAuthorFemale1 0.1010 0.0969 1.04 0.29957
## LastAuthorFemale1 0.1375 0.1091 1.26 0.20970
## UniqueAuthors2 0.3369 0.1486 2.27 0.02502 *
## UniqueAuthors3 0.5355 0.1377 3.89 0.00016 ***
## UniqueAuthors4 0.3064 0.1552 1.97 0.05043 .
## UniqueAuthors5 0.3079 0.1699 1.81 0.07216 .
## Year1997 -0.3029 0.3821 -0.79 0.42944
## Year1998 -0.0837 0.3573 -0.23 0.81508
## Year1999 -0.1273 0.3159 -0.40 0.68755
```

```

## Year2000      -0.4294      0.3406      -1.26      0.20966
## Year2001      0.0467      0.3122      0.15      0.88123
## Year2002     -0.2506      0.3041     -0.82      0.41137
## Year2003     -0.1497      0.3526     -0.42      0.67173
## Year2004      0.0593      0.2759      0.21      0.83014
## Year2005      0.0431      0.3050      0.14      0.88796
## Year2006     -0.0502      0.2894     -0.17      0.86256
## Year2007      0.1196      0.3743      0.32      0.74979
## Year2008      0.0346      0.2807      0.12      0.90199
## Year2009     -0.2119      0.2814     -0.75      0.45271
## Year2010     -0.2590      0.2840     -0.91      0.36350
## Year2011     -0.1994      0.2999     -0.66      0.50735
## Year2012      0.0131      0.3174      0.04      0.96725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.251, Adjusted R-squared:  0.127
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 140 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.261  0.865  0.951  0.901  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.45e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.357 1      1.165
## LastAuthorFemale  1.709 1      1.307
## Year              2.239 16      1.026

```

## Residuals from first and last author



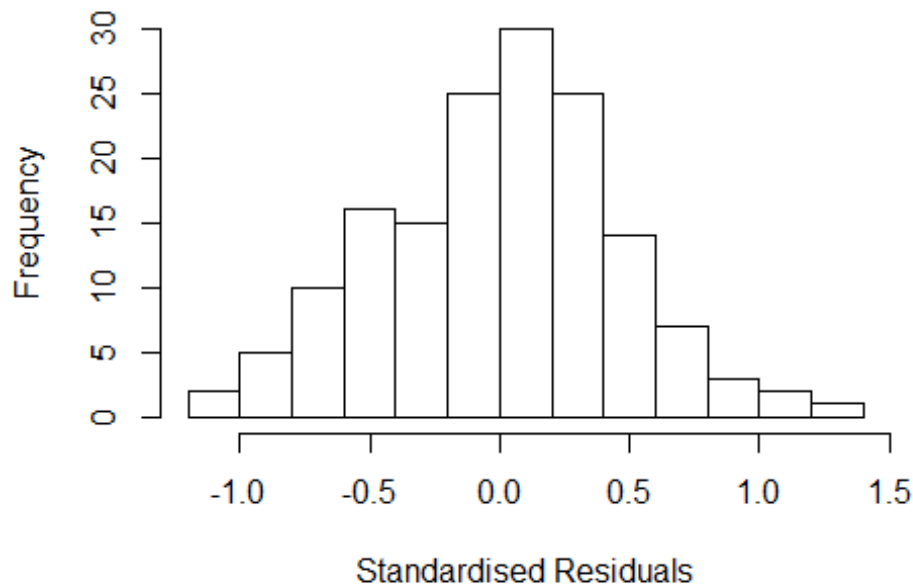
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0564 -0.3112 0.0224 0.3136 1.2987
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0564 0.3496 3.02 0.003 **
## FirstAuthorFemale1 0.1077 0.0952 1.13 0.260
## LastAuthorFemale1 0.1296 0.1019 1.27 0.206
## Year1997 -0.2791 0.4301 -0.65 0.518
## Year1998 -0.0706 0.4051 -0.17 0.862
## Year1999 -0.1382 0.4051 -0.34 0.733
## Year2000 -0.6392 0.4206 -1.52 0.131
## Year2001 -0.0486 0.3773 -0.13 0.898
## Year2002 -0.2738 0.3815 -0.72 0.474
## Year2003 -0.1421 0.4361 -0.33 0.745
## Year2004 -0.0112 0.3709 -0.03 0.976
## Year2005 0.0275 0.3778 0.07 0.942
```

```

## Year2006          -0.1050      0.3668   -0.29    0.775
## Year2007           0.0108      0.3969    0.03    0.978
## Year2008           0.0235      0.3588    0.07    0.948
## Year2009          -0.1371      0.3635   -0.38    0.707
## Year2010          -0.2837      0.3637   -0.78    0.437
## Year2011          -0.1624      0.3719   -0.44    0.663
## Year2012           0.1111      0.4045    0.27    0.784
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.111, Adjusted R-squared:  -0.00637
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 137 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.455  0.883   0.952   0.913   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.45e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.316 1      1.147
## Year              1.316 16      1.009

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0635 -0.3204 0.0486 0.3055 1.3014
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0583 0.3529 3.00 0.0032 **
## FirstAuthorFemale1 0.1294 0.0954 1.36 0.1772
## Year1997 -0.2837 0.4330 -0.66 0.5134
## Year1998 -0.0515 0.4138 -0.12 0.9012
## Year1999 -0.1057 0.4001 -0.26 0.7920
## Year2000 -0.6182 0.4287 -1.44 0.1516
## Year2001 -0.0502 0.3789 -0.13 0.8949
## Year2002 -0.2644 0.3868 -0.68 0.4955
## Year2003 -0.1338 0.4340 -0.31 0.7584
## Year2004 -0.0132 0.3741 -0.04 0.9719
## Year2005 0.0220 0.3817 0.06 0.9541
## Year2006 -0.0852 0.3695 -0.23 0.8180
```

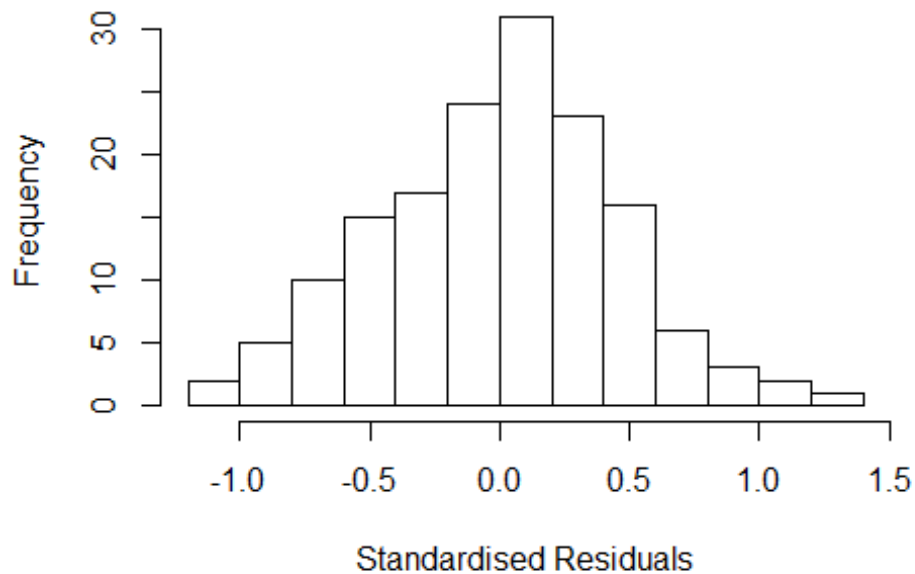
```

## Year2007          0.0509      0.3930      0.13      0.8972
## Year2008          0.0296      0.3617      0.08      0.9348
## Year2009         -0.1276      0.3670     -0.35      0.7287
## Year2010         -0.2393      0.3674     -0.65      0.5160
## Year2011         -0.1242      0.3699     -0.34      0.7375
## Year2012          0.1008      0.4060      0.25      0.8042
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.104, Adjusted R-squared:  -0.0071
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.444  0.886  0.955  0.913  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.45e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.497 1          1.223
## Year            1.497 16          1.013

```



## Residuals from last author



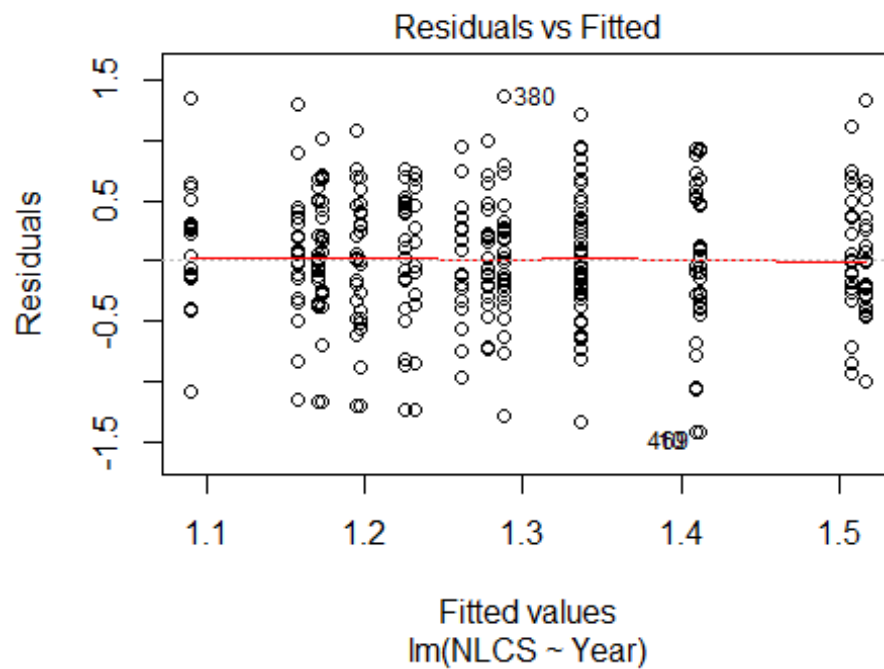
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0576 -0.3038 -0.0018 0.2968 1.3003
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0576 0.3519 3.01 0.0032 **
## LastAuthorFemale1 0.1530 0.1005 1.52 0.1303
## Year1997 -0.2819 0.4329 -0.65 0.5160
## Year1998 -0.0521 0.4091 -0.13 0.8989
## Year1999 -0.1446 0.4077 -0.35 0.7234
## Year2000 -0.6459 0.4223 -1.53 0.1285
## Year2001 -0.0269 0.3777 -0.07 0.9434
## Year2002 -0.2508 0.3825 -0.66 0.5132
## Year2003 -0.1131 0.4313 -0.26 0.7935
## Year2004 -0.0124 0.3731 -0.03 0.9735
## Year2005 0.0442 0.3766 0.12 0.9067
## Year2006 -0.0897 0.3672 -0.24 0.8074
```

```

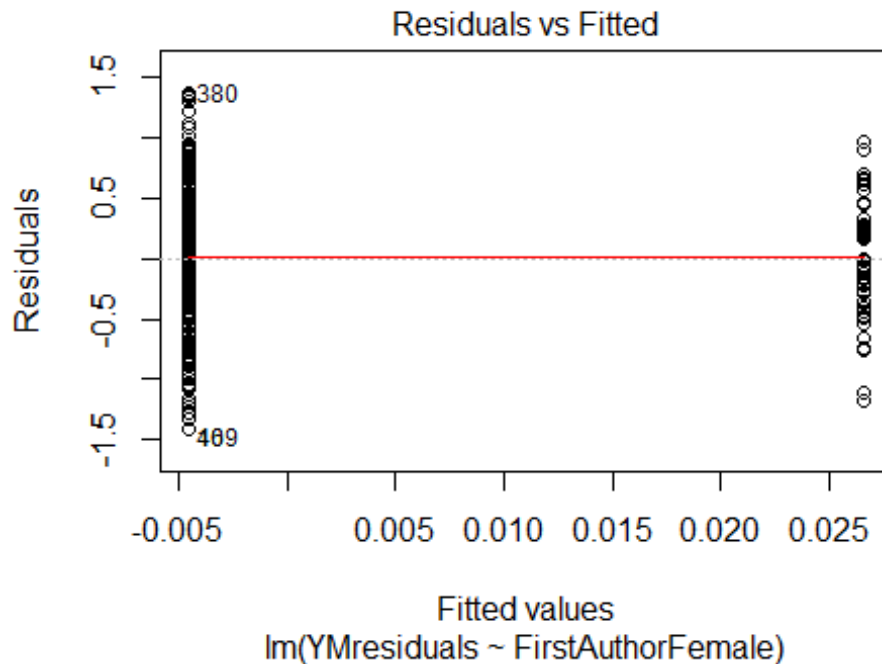
## Year2007          0.0286      0.3933      0.07      0.9422
## Year2008          0.0336      0.3631      0.09      0.9263
## Year2009         -0.0899      0.3652     -0.25      0.8059
## Year2010         -0.2506      0.3644     -0.69      0.4927
## Year2011         -0.1331      0.3785     -0.35      0.7255
## Year2012          0.1519      0.4141      0.37      0.7143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.104, Adjusted R-squared:  -0.00667
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.448  0.884  0.953  0.914  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.45e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 155"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1507"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   94   70  112   92   80   85   91   78   93   84   59   84   72   52   47
## 2011 2012
##   50   60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   21   26   26   19   16   27   18   23   24   20   19   20   22   15
## 2011 2012

```

```
## 19 22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 21 23 25 18 14 25 16 17 15 19 14 19 19 13
## 2011 2012
## 18 15
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 11, df = 16, p-value = 0.8
```

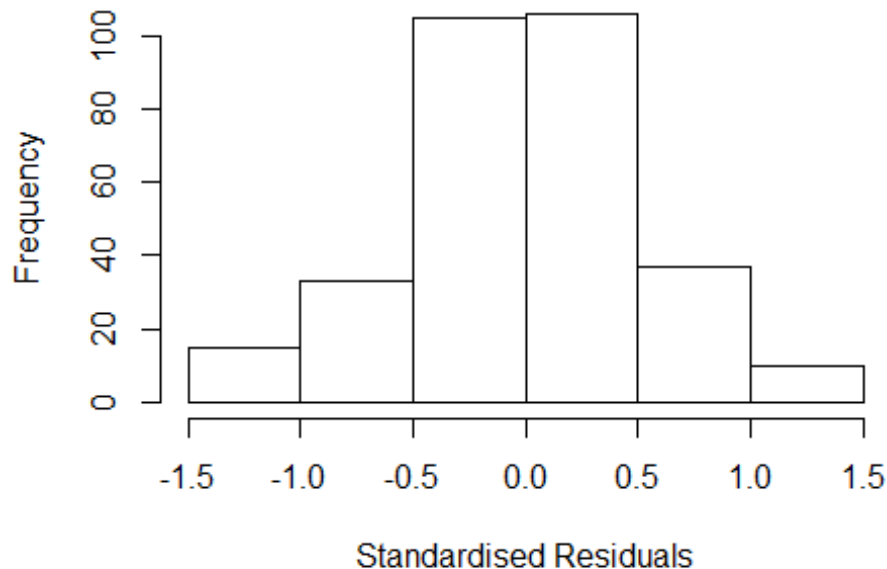


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.2694873523412"
## [1] "Male first author team size 2018 geometric mean: 2.80229967818958"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 580, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.21870785509425"
## [1] "Male last author team size 2018 geometric mean: 2.86352711245201"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 320, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.422 1          1.193
## LastAuthorFemale  1.492 1          1.221
## UniqueAuthors     3.145 4          1.154
## Year              4.047 16          1.045
```

## Residuals from first and last author and team size



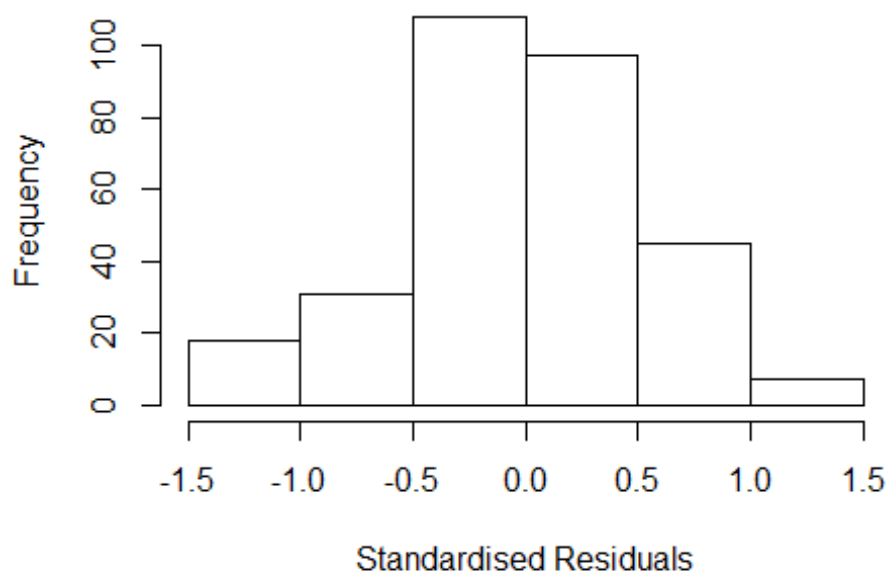
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.49852 -0.30782 -0.00115 0.35407 1.41089
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25171 0.25063 4.99 1e-06 ***
## FirstAuthorFemale1 -0.04876 0.08349 -0.58 0.5596
## LastAuthorFemale1 0.05772 0.08386 0.69 0.4918
## UniqueAuthors2 0.09855 0.08014 1.23 0.2198
## UniqueAuthors3 0.20570 0.10161 2.02 0.0439 *
## UniqueAuthors4 0.43862 0.18231 2.41 0.0168 *
## UniqueAuthors5 0.38608 0.14526 2.66 0.0083 **
## Year1997 -0.05301 0.27114 -0.20 0.8451
## Year1998 0.12097 0.25756 0.47 0.6389
## Year1999 -0.06605 0.25579 -0.26 0.7964
```

```

## Year2000      0.07327    0.26358    0.28    0.7812
## Year2001     -0.17497    0.28049   -0.62    0.5333
## Year2002      0.00121    0.26038    0.00    0.9963
## Year2003     -0.15186    0.26541   -0.57    0.5677
## Year2004     -0.07649    0.31155   -0.25    0.8062
## Year2005     -0.25047    0.28713   -0.87    0.3838
## Year2006     -0.19181    0.27072   -0.71    0.4792
## Year2007     -0.22554    0.29656   -0.76    0.4476
## Year2008     -0.23046    0.24637   -0.94    0.3504
## Year2009      0.08885    0.26198    0.34    0.7348
## Year2010     -0.31533    0.26578   -1.19    0.2365
## Year2011     -0.29066    0.27757   -1.05    0.2959
## Year2012     -0.26984    0.25155   -1.07    0.2843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.0962, Adjusted R-squared:  0.0259
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 279 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.861  0.950  0.888  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.281 1      1.132
## LastAuthorFemale  1.428 1      1.195
## Year              1.692 16      1.017

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4176 -0.3215 -0.0135 0.3655 1.4022
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3750 0.2730 5.04 8.4e-07 ***
## FirstAuthorFemale1 -0.0101 0.0865 -0.12 0.91
## LastAuthorFemale1 0.0532 0.0846 0.63 0.53
## Year1997 -0.1030 0.3038 -0.34 0.73
## Year1998 0.0766 0.2885 0.27 0.79
## Year1999 -0.1292 0.2884 -0.45 0.65
## Year2000 0.0426 0.2999 0.14 0.89
## Year2001 -0.2039 0.3052 -0.67 0.50
## Year2002 -0.0408 0.2934 -0.14 0.89
## Year2003 -0.1532 0.2998 -0.51 0.61
## Year2004 -0.1268 0.3220 -0.39 0.69
## Year2005 -0.2514 0.3111 -0.81 0.42
```

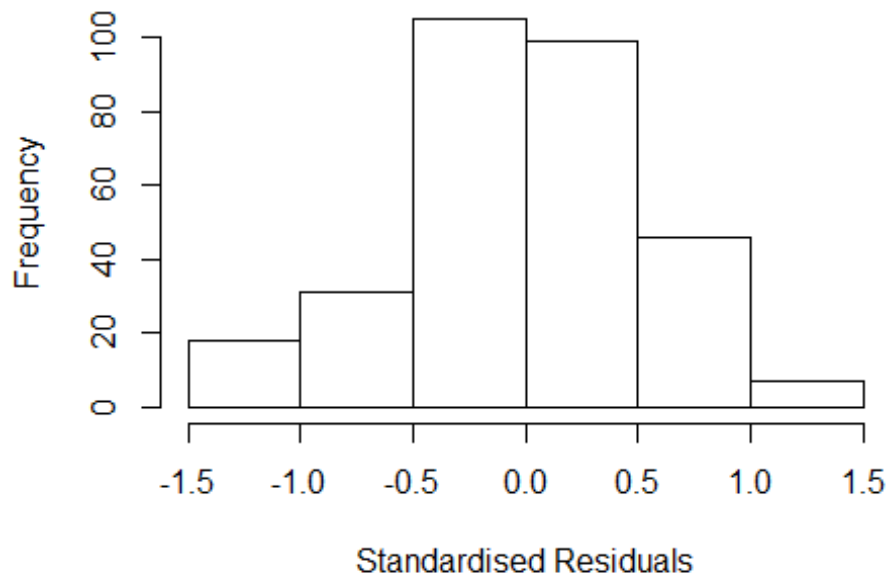
```

## Year2006          -0.2151      0.3102   -0.69      0.49
## Year2007          -0.2051      0.3242   -0.63      0.53
## Year2008          -0.2217      0.2832   -0.78      0.43
## Year2009           0.1073      0.2949    0.36      0.72
## Year2010          -0.2034      0.3114   -0.65      0.51
## Year2011          -0.2871      0.3118   -0.92      0.36
## Year2012          -0.2223      0.2949   -0.75      0.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0469, Adjusted R-squared:  -0.0128
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 279 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.430  0.855   0.951   0.896   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.233 1      1.110
## Year              1.233 16      1.007

```



## Residuals from first author



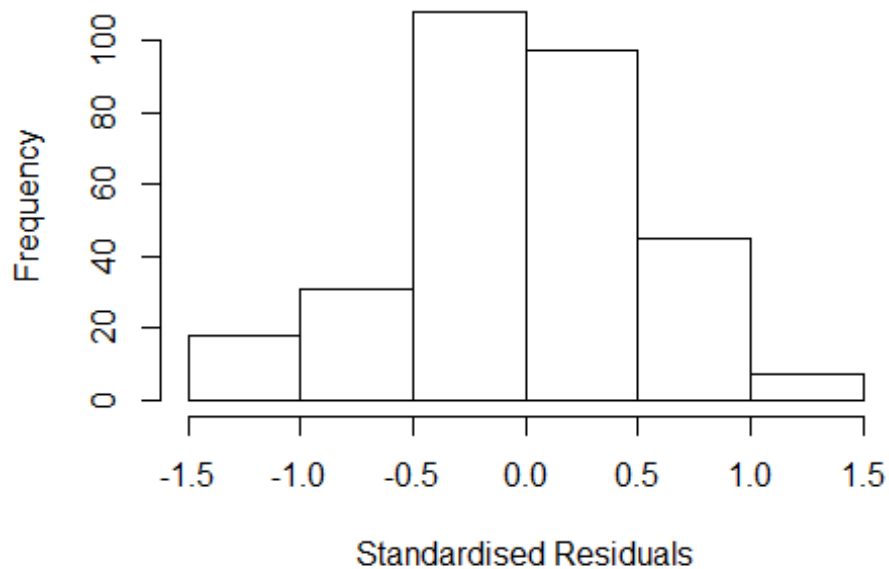
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4206 -0.3248 -0.0142 0.3574 1.3902
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37951 0.28345 4.87 1.9e-06 ***
## FirstAuthorFemale1 0.00413 0.08520 0.05 0.96
## Year1997 -0.10796 0.30984 -0.35 0.73
## Year1998 0.07625 0.30010 0.25 0.80
## Year1999 -0.12172 0.29800 -0.41 0.68
## Year2000 0.04105 0.30878 0.13 0.89
## Year2001 -0.20792 0.31346 -0.66 0.51
## Year2002 -0.03795 0.30273 -0.13 0.90
## Year2003 -0.14634 0.30800 -0.48 0.64
## Year2004 -0.12814 0.33200 -0.39 0.70
## Year2005 -0.25277 0.32096 -0.79 0.43
## Year2006 -0.21007 0.31803 -0.66 0.51
```

```

## Year2007          -0.20868    0.33059   -0.63    0.53
## Year2008          -0.22199    0.29235   -0.76    0.45
## Year2009           0.10152    0.30610    0.33    0.74
## Year2010          -0.20494    0.31778   -0.64    0.52
## Year2011          -0.28608    0.31945   -0.90    0.37
## Year2012          -0.22970    0.30529   -0.75    0.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0463, Adjusted R-squared:  -0.00995
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 281 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.416  0.849   0.950   0.894   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.37 1          1.17
## Year            1.37 16          1.01

```

## Residuals from last author



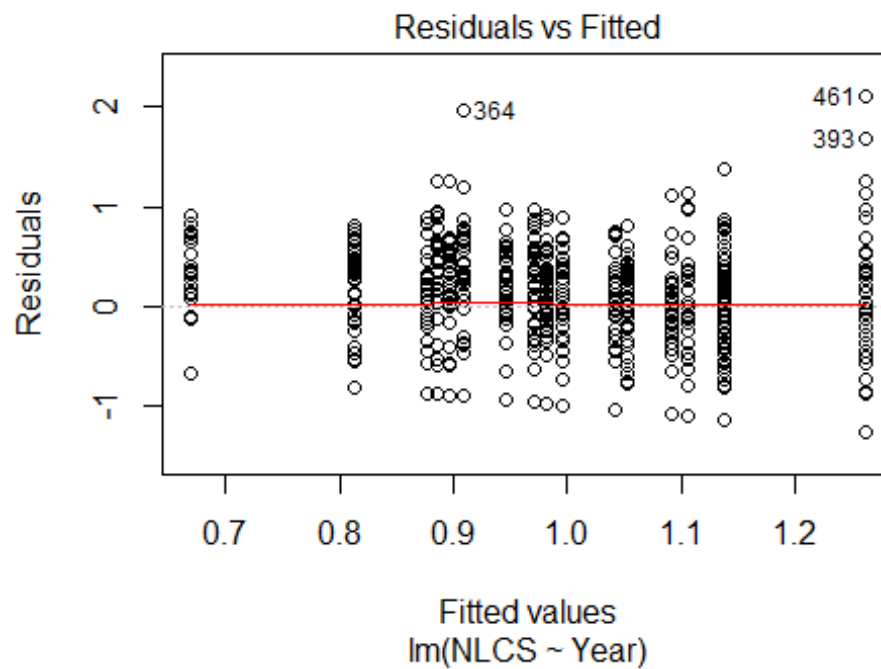
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.417 -0.321 -0.018 0.367 1.402
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3745 0.2750 5.00 1e-06 ***
## LastAuthorFemale1 0.0500 0.0841 0.59 0.55
## Year1997 -0.1030 0.3058 -0.34 0.74
## Year1998 0.0752 0.2922 0.26 0.80
## Year1999 -0.1290 0.2904 -0.44 0.66
## Year2000 0.0426 0.3019 0.14 0.89
## Year2001 -0.2032 0.3066 -0.66 0.51
## Year2002 -0.0423 0.2971 -0.14 0.89
## Year2003 -0.1527 0.3013 -0.51 0.61
## Year2004 -0.1273 0.3252 -0.39 0.70
## Year2005 -0.2515 0.3131 -0.80 0.42
## Year2006 -0.2171 0.3131 -0.69 0.49
```

```

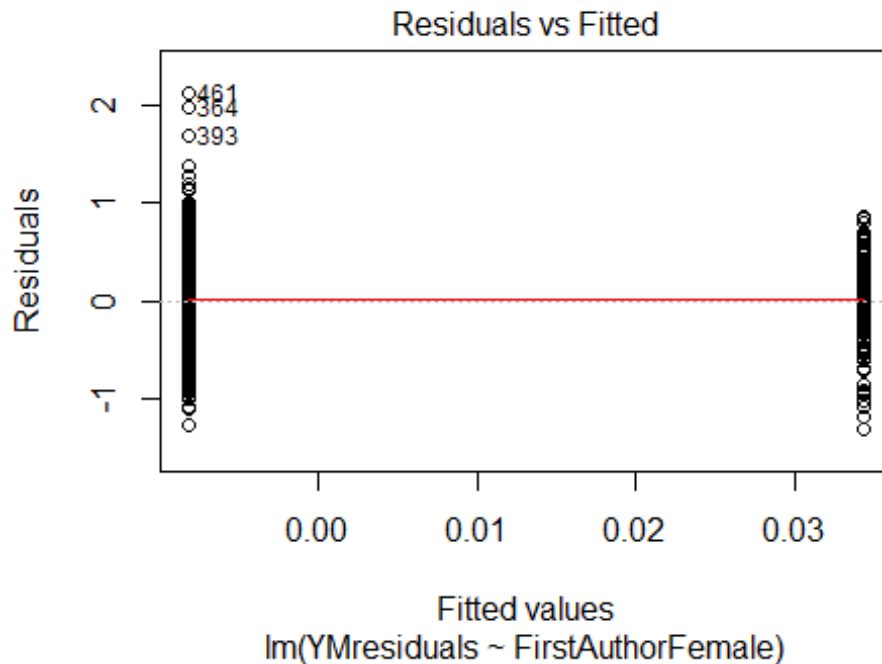
## Year2007          -0.2069      0.3260    -0.63      0.53
## Year2008          -0.2225      0.2862    -0.78      0.44
## Year2009           0.1049      0.2991     0.35      0.73
## Year2010          -0.2023      0.3125    -0.65      0.52
## Year2011          -0.2872      0.3138    -0.92      0.36
## Year2012          -0.2239      0.2986    -0.75      0.45
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.0469, Adjusted R-squared:  -0.00935
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.423  0.854  0.951  0.895  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 306"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1508"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   77   96   84   77  100   82   89   80   96   56   63   70   79   57   56
## 2011 2012
##   58   56
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   35   52   38   50   52   45   48   43   50   32   40   39   43   33   26
## 2011 2012

```

```
## 30 37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 49 35 45 49 43 42 39 47 26 36 31 35 28 24
## 2011 2012
## 25 32
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 58, df = 16, p-value = 1e-06
```

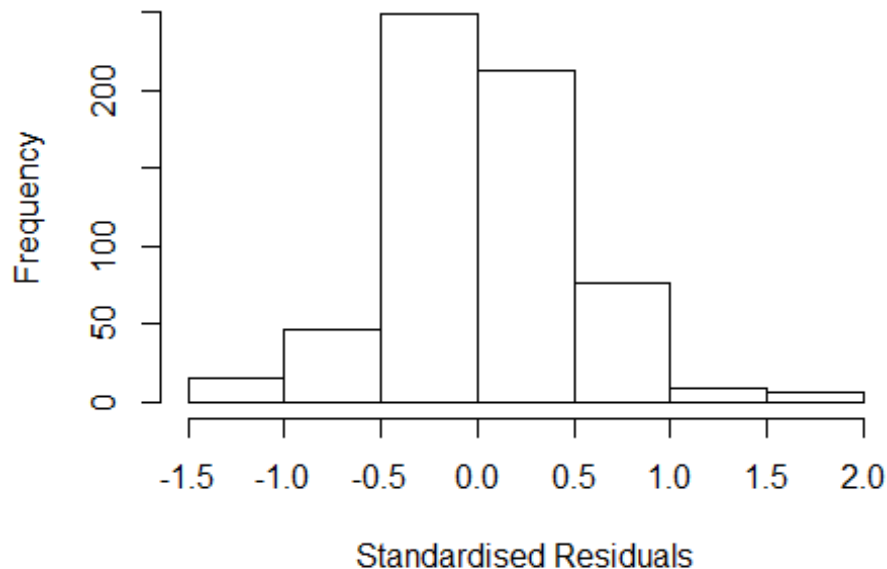


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.3, df = 1, p-value = 0.007
```



```
## [1] "Female first author team size 2018 geometric mean: 3.38980204418032"
## [1] "Male first author team size 2018 geometric mean: 3.28326832771534"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.0864343996059"
## [1] "Male last author team size 2018 geometric mean: 3.23850133581854"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 410, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.259 1          1.122
## LastAuthorFemale  1.200 1          1.096
## UniqueAuthors    1.744 4          1.072
## Year              2.501 16          1.029
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40318 -0.29469 -0.00906 0.29232 1.94456
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2732 0.1120 2.44 0.0150 *
## FirstAuthorFemale1 0.0395 0.0483 0.82 0.4145
## LastAuthorFemale1 -0.0690 0.0544 -1.27 0.2051
## UniqueAuthors2 0.6327 0.0862 7.34 7.1e-13 ***
## UniqueAuthors3 0.8130 0.0658 12.35 < 2e-16 ***
## UniqueAuthors4 0.8845 0.0683 12.96 < 2e-16 ***
## UniqueAuthors5 0.8443 0.0587 14.39 < 2e-16 ***
## Year1997 0.1087 0.1373 0.79 0.4288
## Year1998 -0.0511 0.1342 -0.38 0.7034
## Year1999 0.0214 0.1274 0.17 0.8664
```

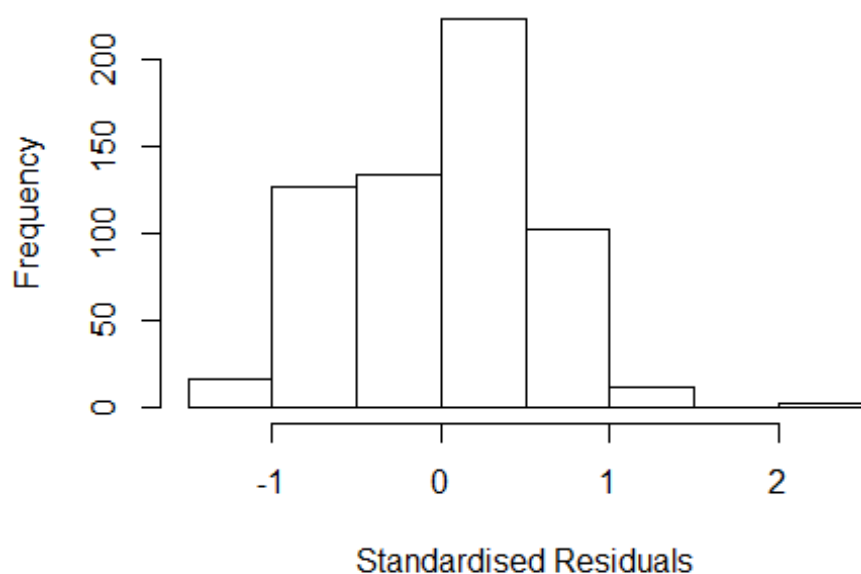
```

## Year2000          0.4577      0.1498      3.06      0.0023 **
## Year2001          0.3197      0.1333      2.40      0.0168 *
## Year2002          0.0745      0.1493      0.50      0.6179
## Year2003          0.0359      0.1294      0.28      0.7814
## Year2004          0.0653      0.1215      0.54      0.5914
## Year2005          0.0973      0.1467      0.66      0.5076
## Year2006          0.1514      0.1258      1.20      0.2291
## Year2007          0.1010      0.1359      0.74      0.4576
## Year2008          0.0119      0.1249      0.09      0.9244
## Year2009         -0.0282      0.1443     -0.20      0.8451
## Year2010          0.2191      0.1385      1.58      0.1143
## Year2011          0.0245      0.1402      0.17      0.8613
## Year2012         -0.1011      0.1341     -0.75      0.4515
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.398
## Multiple R-squared:  0.432, Adjusted R-squared:  0.411
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 3 observations c(21,157,209) are outliers with |weight| = 0 ( < 0.00016);
## 47 weights are ~= 1. The remaining 566 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8360 0.9450 0.8750 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.62e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.248 1          1.117
## LastAuthorFemale 1.106 1          1.052
## Year          1.360 16          1.010

```



## Residuals from first and last author



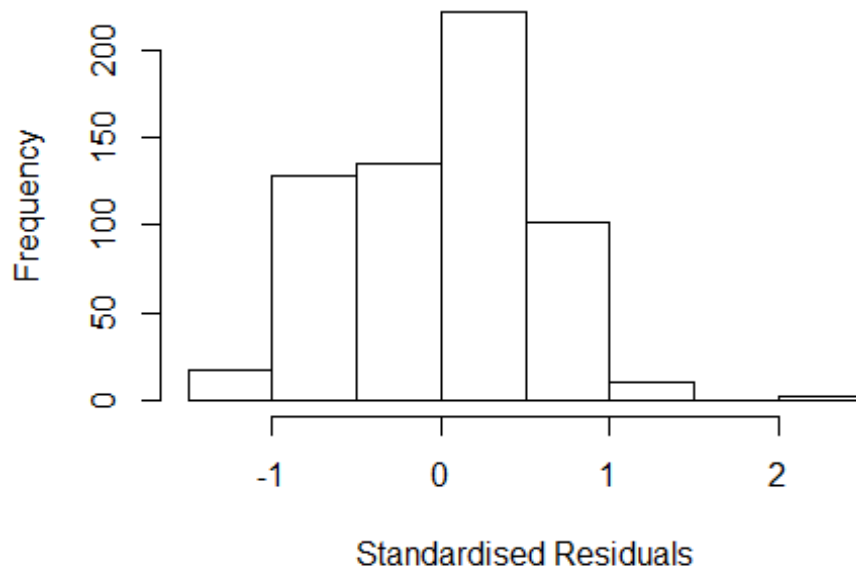
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3203 -0.4407 0.0662 0.3749 2.1314
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7923 0.1673 4.73 2.7e-06 ***
## FirstAuthorFemale1 0.0817 0.0585 1.40 0.163
## LastAuthorFemale1 -0.1064 0.0679 -1.57 0.118
## Year1997 0.0309 0.1950 0.16 0.874
## Year1998 -0.1579 0.2099 -0.75 0.452
## Year1999 0.0342 0.2125 0.16 0.872
## Year2000 0.4463 0.2173 2.05 0.040 *
## Year2001 0.3341 0.1837 1.82 0.069 .
## Year2002 0.1501 0.2000 0.75 0.453
## Year2003 0.1025 0.2042 0.50 0.616
## Year2004 0.2047 0.1839 1.11 0.266
## Year2005 0.2585 0.2028 1.27 0.203
```

```

## Year2006          0.2618      0.1812      1.44      0.149
## Year2007          0.2893      0.1815      1.59      0.111
## Year2008          0.1445      0.1893      0.76      0.446
## Year2009          0.2319      0.1908      1.22      0.225
## Year2010          0.3917      0.1935      2.02      0.043 *
## Year2011          0.2332      0.1853      1.26      0.209
## Year2012          0.0747      0.1918      0.39      0.697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0686, Adjusted R-squared:  0.0406
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 569 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0752 0.8280 0.9390 0.8950 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.231 1      1.109
## Year              1.231 16      1.007

```

## Residuals from first author



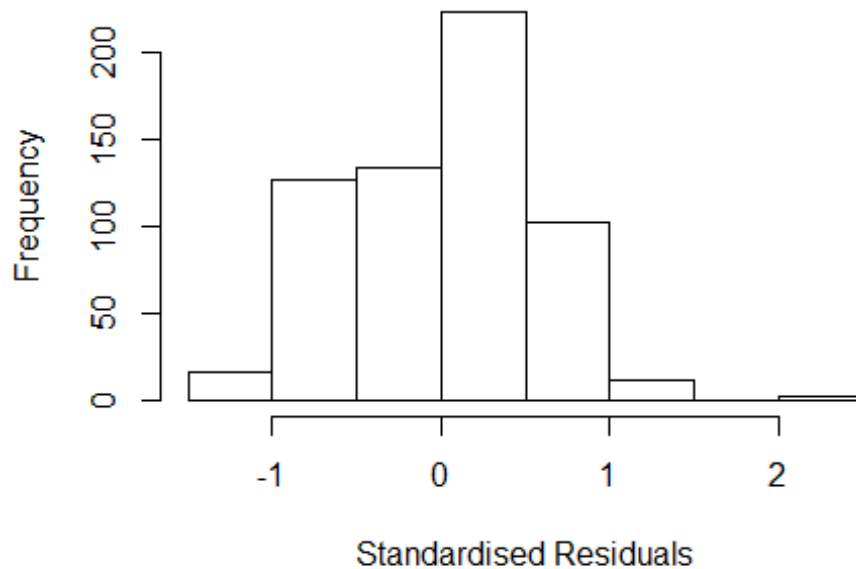
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3065 -0.4229 0.0666 0.3760 2.1356
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7909 0.1687 4.69 3.4e-06 ***
## FirstAuthorFemale1 0.0722 0.0611 1.18 0.238
## Year1997 0.0258 0.1959 0.13 0.895
## Year1998 -0.1676 0.2122 -0.79 0.430
## Year1999 0.0244 0.2122 0.11 0.909
## Year2000 0.4434 0.2174 2.04 0.042 *
## Year2001 0.3197 0.1857 1.72 0.086 .
## Year2002 0.1418 0.2004 0.71 0.480
## Year2003 0.0911 0.2051 0.44 0.657
## Year2004 0.2016 0.1853 1.09 0.277
## Year2005 0.2415 0.2024 1.19 0.233
## Year2006 0.2356 0.1819 1.30 0.196
```

```

## Year2007          0.2816      0.1823      1.54      0.123
## Year2008          0.1359      0.1902      0.71      0.475
## Year2009          0.2003      0.1930      1.04      0.300
## Year2010          0.3706      0.1934      1.92      0.056 .
## Year2011          0.2145      0.1863      1.15      0.250
## Year2012          0.0582      0.1927      0.30      0.763
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0652, Adjusted R-squared:  0.0386
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 573 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0751 0.8290 0.9420 0.8960 0.9790 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.083 1      1.041
## Year              1.083 16      1.002

```

## Residuals from last author



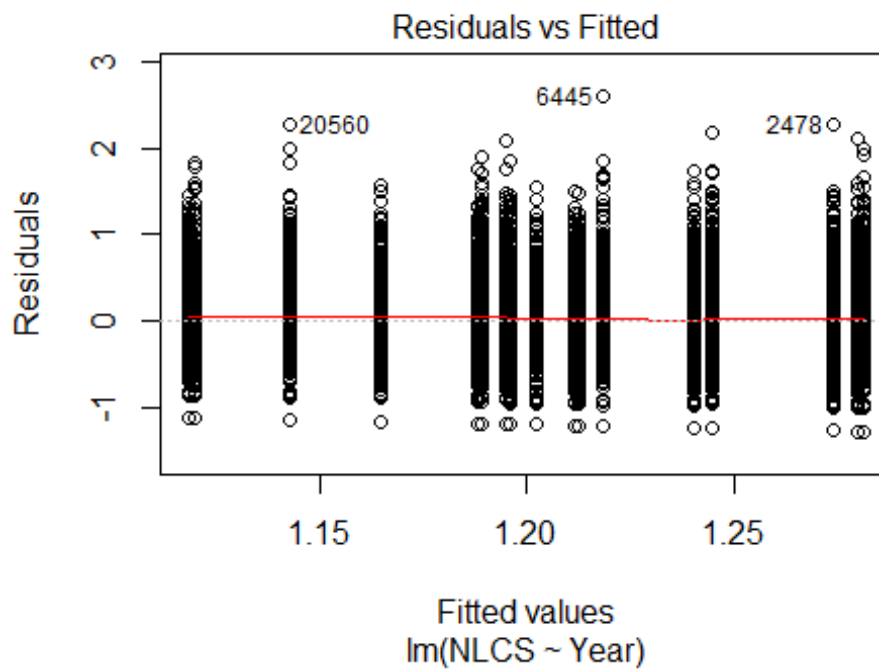
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2491 -0.4372 0.0596 0.3693 2.1209
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8021 0.1659 4.84 1.7e-06 ***
## LastAuthorFemale1 -0.0958 0.0665 -1.44 0.150
## Year1997 0.0220 0.1945 0.11 0.910
## Year1998 -0.1562 0.2094 -0.75 0.456
## Year1999 0.0332 0.2125 0.16 0.876
## Year2000 0.4470 0.2184 2.05 0.041 *
## Year2001 0.3439 0.1815 1.89 0.059 .
## Year2002 0.1642 0.1993 0.82 0.410
## Year2003 0.1246 0.2049 0.61 0.543
## Year2004 0.2229 0.1835 1.21 0.225
## Year2005 0.2612 0.2032 1.29 0.199
## Year2006 0.2687 0.1804 1.49 0.137
```

```

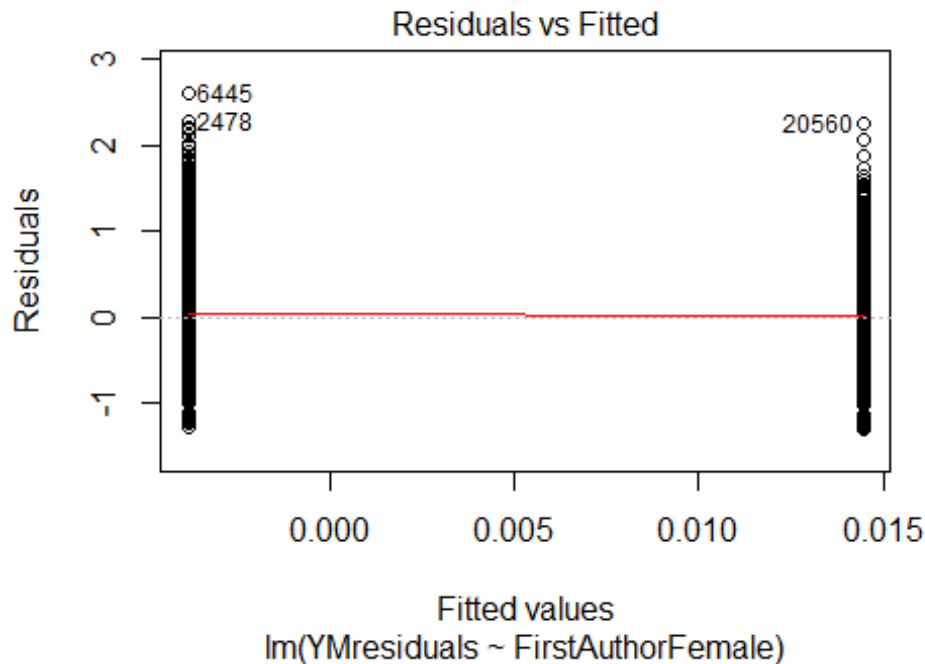
## Year2007          0.2918      0.1810      1.61      0.107
## Year2008          0.1480      0.1897      0.78      0.436
## Year2009          0.2351      0.1915      1.23      0.220
## Year2010          0.3897      0.1924      2.03      0.043 *
## Year2011          0.2396      0.1854      1.29      0.197
## Year2012          0.0727      0.1907      0.38      0.703
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0663, Adjusted R-squared:  0.0397
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 49 weights are ~= 1. The remaining 567 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0792 0.8320 0.9400 0.8940 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 616"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1600"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1668 1638 1652 1454 1716 1560 1525 1230 1264 1175 1382 1346 1338 1517 1415
## 2011 2012
## 1527 1436
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 995 1011 1014 902 881 567 994 801 815 762 866 901 857 947 896
## 2011 2012

```

```
## 972 925
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 859 875 901 809 789 506 886 722 740 660 760 800 775 842 800
## 2011 2012
## 860 821
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```



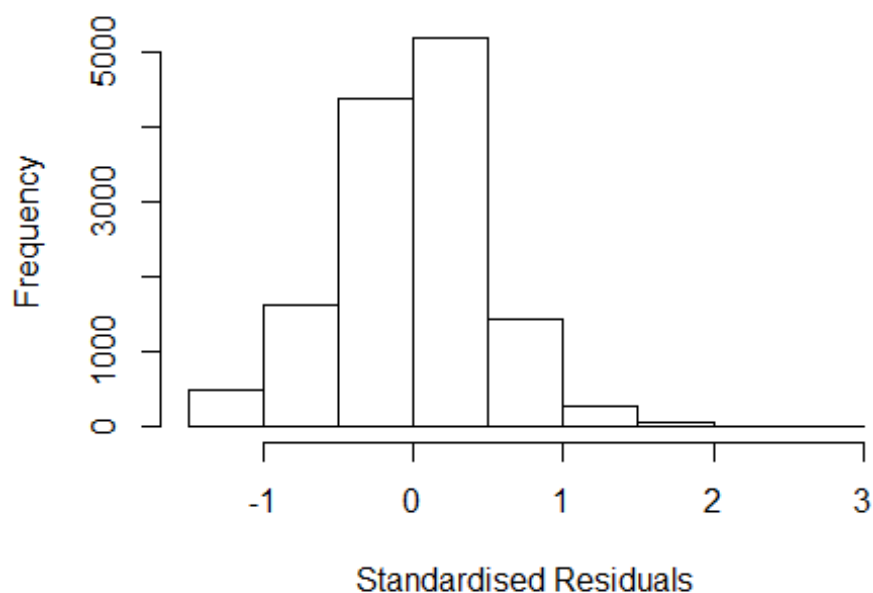
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 5.11927124332424"
## [1] "Male first author team size 2018 geometric mean: 4.17771693057365"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 87000, p-value = 4e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.95729175955355"
## [1] "Male last author team size 2018 geometric mean: 4.31741175406262"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 59000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1 1.017
## LastAuthorFemale 1.006 1 1.003
## UniqueAuthors 1.051 4 1.006
## Year 1.071 16 1.002
```



## Residuals from first and last author and team size



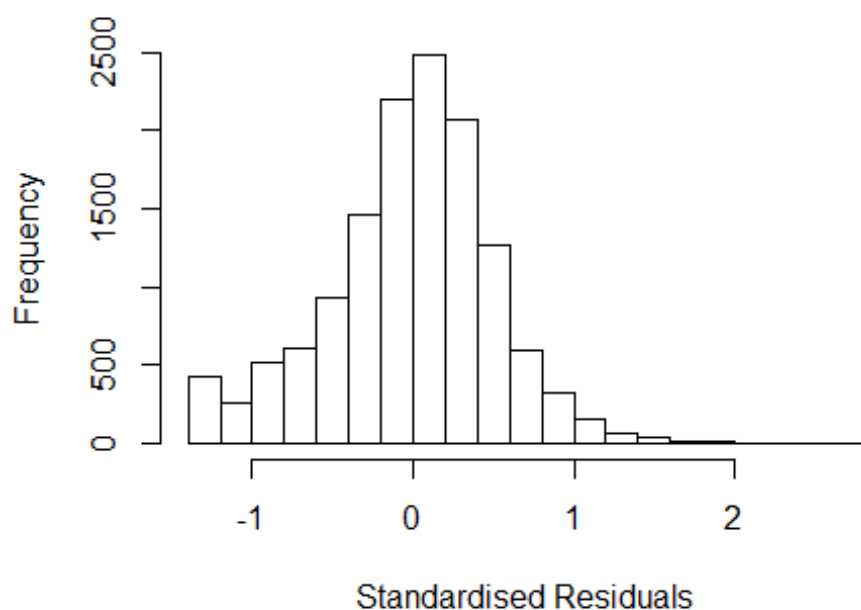
```
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 716      0001094662 3.434 1996      1600      1      2.561
## 6445     0347417134 3.818 1999      1600      1      2.954
## 20560    34249889935 3.412 2007      1600      3      2.611
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3661 -0.2914  0.0209  0.3039  2.9538
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.87258    0.02826   30.88 < 2e-16 ***
## FirstAuthorFemale1 -0.01061    0.01031   -1.03  0.30329
## LastAuthorFemale1  0.00281    0.01285    0.22  0.82674
## UniqueAuthors2     0.34285    0.02627   13.05 < 2e-16 ***
## UniqueAuthors3     0.35933    0.02553   14.08 < 2e-16 ***
## UniqueAuthors4     0.39597    0.02560   15.47 < 2e-16 ***
## UniqueAuthors5     0.43408    0.02486   17.46 < 2e-16 ***
## Year1997          0.04294    0.02296    1.87  0.06147 .
```

```

## Year1998          0.07006      0.02296      3.05  0.00228 **
## Year1999          -0.00837      0.02311     -0.36  0.71731
## Year2000          -0.01400      0.02339     -0.60  0.54965
## Year2001           0.00882      0.02794      0.32  0.75239
## Year2002          -0.03329      0.02230     -1.49  0.13550
## Year2003          -0.09652      0.02619     -3.69  0.00023 ***
## Year2004          -0.05306      0.02514     -2.11  0.03481 *
## Year2005          -0.00929      0.02461     -0.38  0.70590
## Year2006          -0.06752      0.02625     -2.57  0.01012 *
## Year2007          -0.06404      0.02690     -2.38  0.01728 *
## Year2008          -0.01708      0.02493     -0.69  0.49322
## Year2009          -0.01514      0.02489     -0.61  0.54290
## Year2010           0.00472      0.02268      0.21  0.83500
## Year2011           0.05204      0.02215      2.35  0.01879 *
## Year2012          -0.01132      0.02245     -0.50  0.61399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.0631, Adjusted R-squared:  0.0616
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 10 observations c(355,376,617,1117,2470,2947,5841,8884,10011,11838)
## are outliers with |weight| = 0 ( < 7.5e-06);
## 1163 weights are ~= 1. The remaining 12232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0027 0.8500  0.9500  0.8840  0.9860  0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.46e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1          1.013
## LastAuthorFemale  1.005 1          1.002
## Year              1.030 16          1.001

```

## Residuals from first and last author



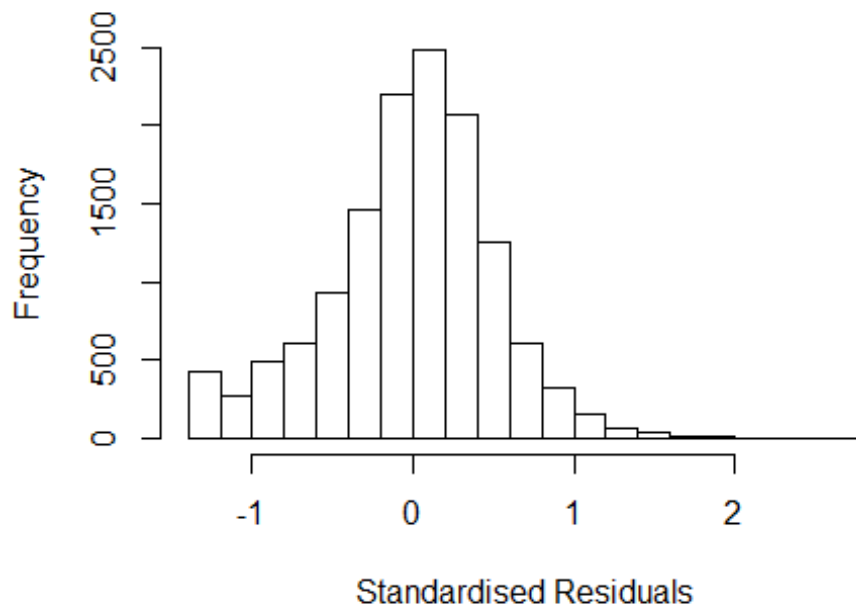
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6445 0347417134 3.818 1999      1600      1      2.611
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3210 -0.3028  0.0271  0.3073  2.6110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22456    0.01632   75.04 < 2e-16 ***
## FirstAuthorFemale1 0.01194    0.01046    1.14  0.25381
## LastAuthorFemale1 0.01452    0.01306    1.11  0.26632
## Year1997         0.04222    0.02324    1.82  0.06932 .
## Year1998         0.07002    0.02320    3.02  0.00255 **
## Year1999        -0.01756    0.02347   -0.75  0.45442
## Year2000        -0.00810    0.02386   -0.34  0.73406
## Year2001         0.00183    0.02840    0.06  0.94871
## Year2002        -0.02202    0.02288   -0.96  0.33592
## Year2003        -0.09764    0.02694   -3.62  0.00029 ***
## Year2004        -0.05056    0.02532   -2.00  0.04582 *
## Year2005         0.00858    0.02470    0.35  0.72837
```

```

## Year2006          -0.05681      0.02676      -2.12  0.03378 *
## Year2007          -0.05630      0.02760      -2.04  0.04142 *
## Year2008          -0.00205      0.02559      -0.08  0.93602
## Year2009          -0.00411      0.02511      -0.16  0.87008
## Year2010           0.01161      0.02339       0.50  0.61980
## Year2011           0.06802      0.02238       3.04  0.00238 **
## Year2012           0.00347      0.02277       0.15  0.87878
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.00771,    Adjusted R-squared:  0.00638
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 5 observations c(355,617,1117,2947,8884)
## are outliers with |weight| = 0 ( < 7.5e-06);
## 1157 weights are ~ = 1. The remaining 12243 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000  0.849  0.949   0.882   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.46e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1      1.013
## Year              1.026 16      1.001

```

## Residuals from first author

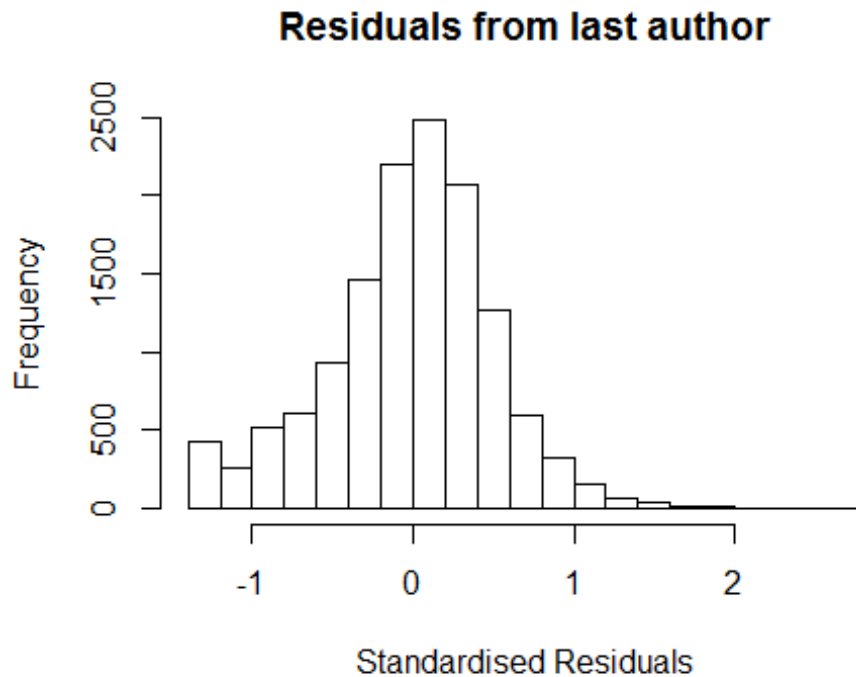


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6445 0347417134 3.818 1999      1600      1      2.611
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3086 -0.3036  0.0267  0.3077  2.6093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22651    0.01618   75.78 < 2e-16 ***
## FirstAuthorFemale1 0.01247    0.01048    1.19  0.23436
## Year1997        0.04170    0.02323    1.80  0.07260 .
## Year1998        0.06958    0.02320    3.00  0.00271 **
## Year1999       -0.01780    0.02348   -0.76  0.44858
## Year2000       -0.00814    0.02387   -0.34  0.73303
## Year2001        0.00146    0.02840    0.05  0.95899
## Year2002       -0.02234    0.02288   -0.98  0.32895
## Year2003       -0.09801    0.02693   -3.64  0.00027 ***
## Year2004       -0.05072    0.02533   -2.00  0.04529 *
## Year2005        0.00843    0.02471    0.34  0.73284
## Year2006       -0.05691    0.02676   -2.13  0.03343 *
```

```

## Year2007          -0.05694      0.02761      -2.06      0.03917 *
## Year2008          -0.00224      0.02559      -0.09      0.93030
## Year2009          -0.00438      0.02510      -0.17      0.86163
## Year2010           0.01123      0.02339       0.48      0.63108
## Year2011           0.06775      0.02239       3.03      0.00248 **
## Year2012           0.00332      0.02276       0.15      0.88410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.00761,    Adjusted R-squared:  0.00635
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 5 observations c(355,617,1117,2947,8884)
## are outliers with |weight| = 0 ( < 7.5e-06);
## 1147 weights are ~= 1. The remaining 12253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  0.850  0.949   0.882   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.46e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.004 1      1.002
## Year              1.004 16      1.000

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 6445 0347417134 3.818 1999      1600      1      2.611
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3116 -0.3029  0.0262  0.3079  2.6092
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.226113   0.016264   75.39 < 2e-16 ***
## LastAuthorFemale1 0.015185   0.013079    1.16  0.24568
## Year1997        0.042116   0.023239    1.81  0.06997 .
## Year1998        0.070295   0.023189    3.03  0.00244 **
## Year1999       -0.017362   0.023469   -0.74  0.45944
## Year2000       -0.007503   0.023847   -0.31  0.75304
## Year2001        0.002713   0.028366    0.10  0.92381
## Year2002       -0.020958   0.022827   -0.92  0.35857
## Year2003       -0.096575   0.026877   -3.59  0.00033 ***
## Year2004       -0.049676   0.025290   -1.96  0.04952 *
## Year2005        0.009584   0.024662    0.39  0.69755
## Year2006       -0.055488   0.026721   -2.08  0.03786 *
```

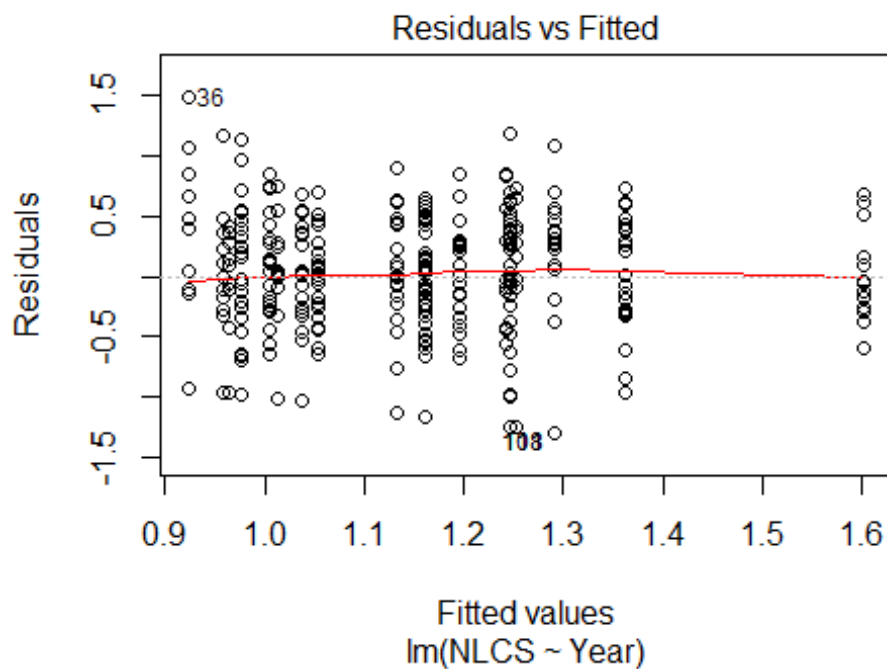
```

## Year2007          -0.054733    0.027537    -1.99    0.04687 *
## Year2008          -0.000442    0.025535    -0.02    0.98620
## Year2009          -0.002670    0.025060    -0.11    0.91515
## Year2010           0.013050    0.023322     0.56    0.57580
## Year2011           0.069661    0.022282     3.13    0.00177 **
## Year2012           0.005239    0.022662     0.23    0.81718
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.0076, Adjusted R-squared:  0.00634
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 5 observations c(355,617,1117,2947,8884)
## are outliers with |weight| = 0 ( < 7.5e-06);
## 1148 weights are ~ = 1. The remaining 12252 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8490 0.9490 0.8820 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.46e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 13405"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1601"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   60   77   62   30   62   52   51   70   49   54   54   64   34   20   22
## 2011 2012
##   34   31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```



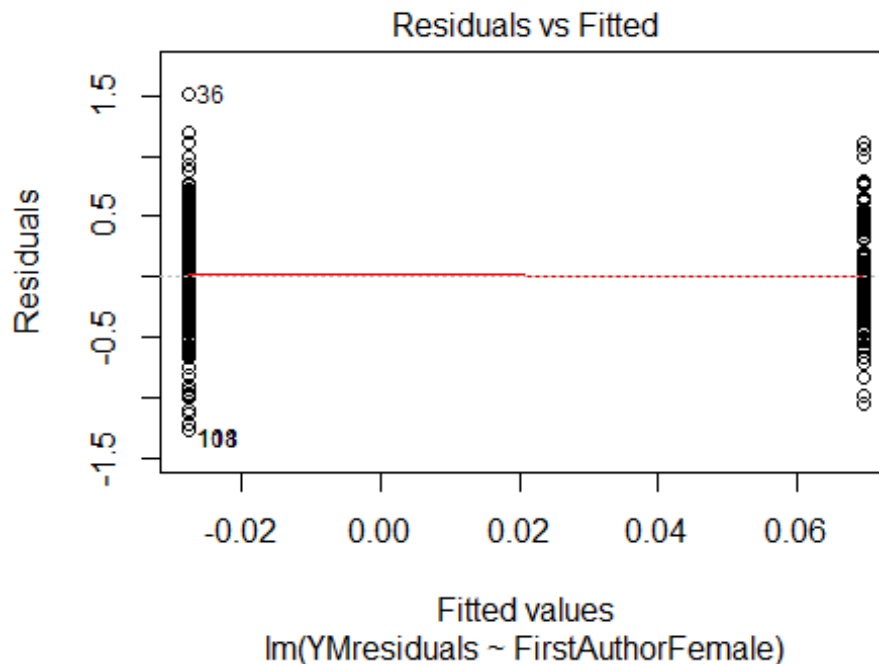
```
## 15 20 10 14 28 12 13 27 19 33 25 40 20 13 12
## 2011 2012
## 25 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 16 8 14 22 12 9 25 16 27 21 35 18 11 12
## 2011 2012
## 24 19
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00041, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 3.54687734174924"
## [1] "Male first author team size 2018 geometric mean: 2.33537867853434"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 420, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.3096617104661"
## [1] "Male last author team size 2018 geometric mean: 3.06146559854297"

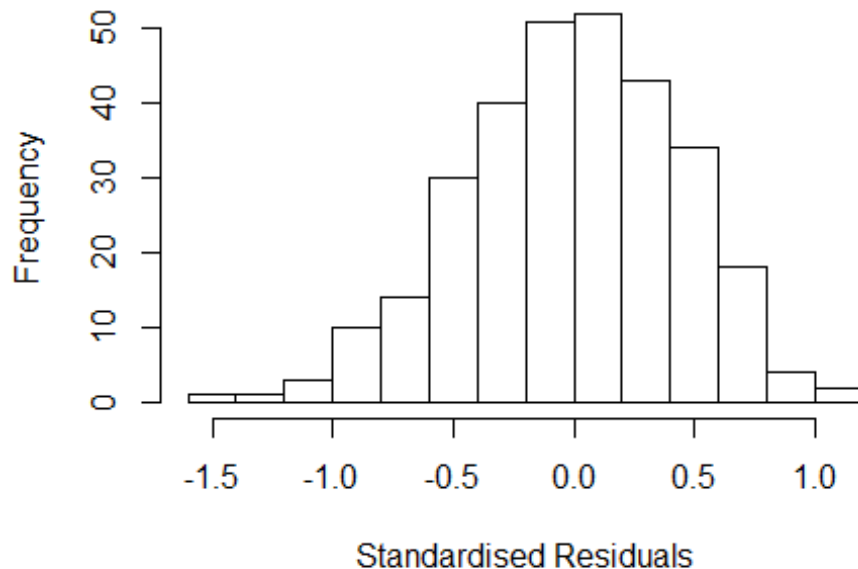
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 150, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.570	1	1.253
LastAuthorFemale	1.747	1	1.322
UniqueAuthors	3.609	4	1.174
Year	5.228	16	1.053

## Residuals from first and last author and team size



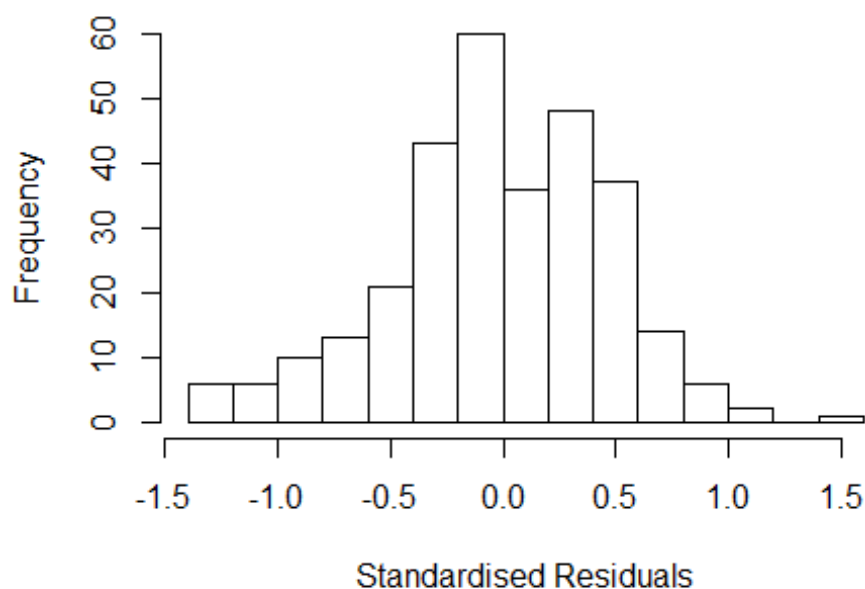
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.48366 -0.28872 0.00157 0.29855 1.19677
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.77481 0.21096 3.67 0.00029 ***
## FirstAuthorFemale1 0.09670 0.06975 1.39 0.16675
## LastAuthorFemale1 0.02034 0.07455 0.27 0.78521
## UniqueAuthors2 0.48381 0.10492 4.61 6.1e-06 ***
## UniqueAuthors3 0.43442 0.09892 4.39 1.6e-05 ***
## UniqueAuthors4 0.32265 0.10420 3.10 0.00216 **
## UniqueAuthors5 0.50083 0.09754 5.13 5.3e-07 ***
## Year1997 0.22504 0.29213 0.77 0.44174
## Year1998 0.09742 0.33338 0.29 0.77033
## Year1999 0.38684 0.22679 1.71 0.08917 .
```

```

## Year2000      -0.16477      0.24478      -0.67      0.50141
## Year2001      -0.24002      0.22772      -1.05      0.29278
## Year2002      -0.37519      0.27682      -1.36      0.17640
## Year2003       0.21581      0.22932       0.94      0.34748
## Year2004      -0.06593      0.25021      -0.26      0.79235
## Year2005       0.07508      0.23261       0.32      0.74711
## Year2006      -0.07118      0.23940      -0.30      0.76643
## Year2007      -0.04606      0.22437      -0.21      0.83750
## Year2008      -0.00600      0.22524      -0.03      0.97878
## Year2009       0.00873      0.27605       0.03      0.97478
## Year2010       0.14000      0.25176       0.56      0.57860
## Year2011      -0.11446      0.22534      -0.51      0.61189
## Year2012       0.06328      0.25212       0.25      0.80201
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.234, Adjusted R-squared:  0.173
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 279 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.875   0.957   0.914   0.989   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.30e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.389 1          1.178
## LastAuthorFemale  1.427 1          1.195
## Year              1.641 16          1.016

```

## Residuals from first and last author



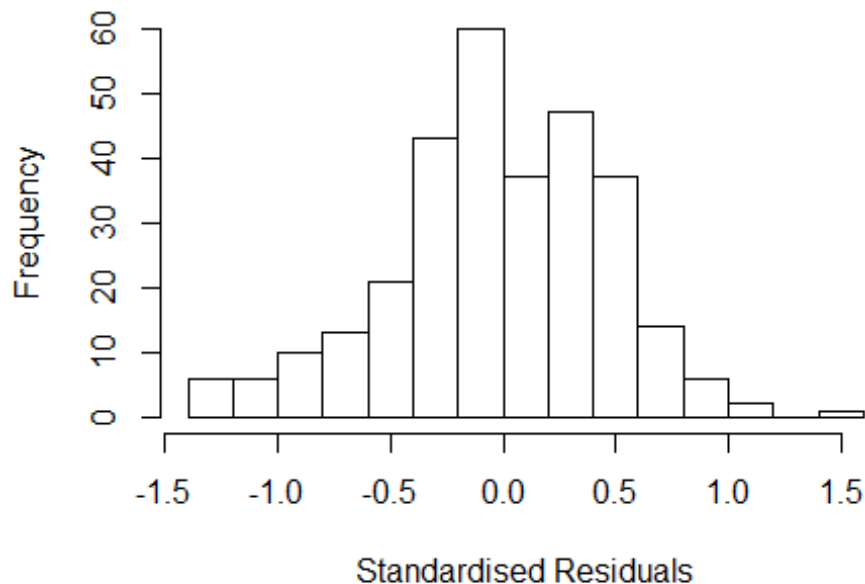
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3228 -0.3028 -0.0264 0.3417 1.4470
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95898 0.30151 3.18 0.0016 **
## FirstAuthorFemale1 0.11569 0.07306 1.58 0.1144
## LastAuthorFemale1 -0.01247 0.07517 -0.17 0.8683
## Year1997 0.36379 0.36116 1.01 0.3147
## Year1998 0.30860 0.41009 0.75 0.4524
## Year1999 0.59724 0.32141 1.86 0.0642 .
## Year2000 -0.00157 0.32639 0.00 0.9962
## Year2001 0.01627 0.32060 0.05 0.9596
## Year2002 -0.14053 0.34000 -0.41 0.6797
## Year2003 0.43782 0.31691 1.38 0.1682
## Year2004 0.15388 0.32481 0.47 0.6360
## Year2005 0.33519 0.31765 1.06 0.2922
```

```

## Year2006          0.09225    0.31879    0.29    0.7725
## Year2007          0.15671    0.31356    0.50    0.6176
## Year2008          0.14067    0.32062    0.44    0.6612
## Year2009          0.00566    0.35144    0.02    0.9872
## Year2010          0.20960    0.34237    0.61    0.5409
## Year2011          0.01515    0.32127    0.05    0.9624
## Year2012          0.20618    0.33743    0.61    0.5417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.48
## Multiple R-squared:  0.123, Adjusted R-squared:  0.0676
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.343  0.884  0.953  0.906  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.262 1      1.123
## Year              1.262 16      1.007

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3254 -0.2982 -0.0272 0.3453 1.4468
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95916 0.30334 3.16 0.0017 **
## FirstAuthorFemale1 0.11246 0.06983 1.61 0.1084
## Year1997 0.36620 0.35779 1.02 0.3069
## Year1998 0.30623 0.40590 0.75 0.4512
## Year1999 0.59552 0.32278 1.84 0.0661 .
## Year2000 -0.00389 0.32710 -0.01 0.9905
## Year2001 0.01564 0.32125 0.05 0.9612
## Year2002 -0.14153 0.34150 -0.41 0.6789
## Year2003 0.43399 0.31592 1.37 0.1706
## Year2004 0.15014 0.32300 0.46 0.6424
## Year2005 0.33419 0.31906 1.05 0.2958
## Year2006 0.09251 0.32043 0.29 0.7730
```

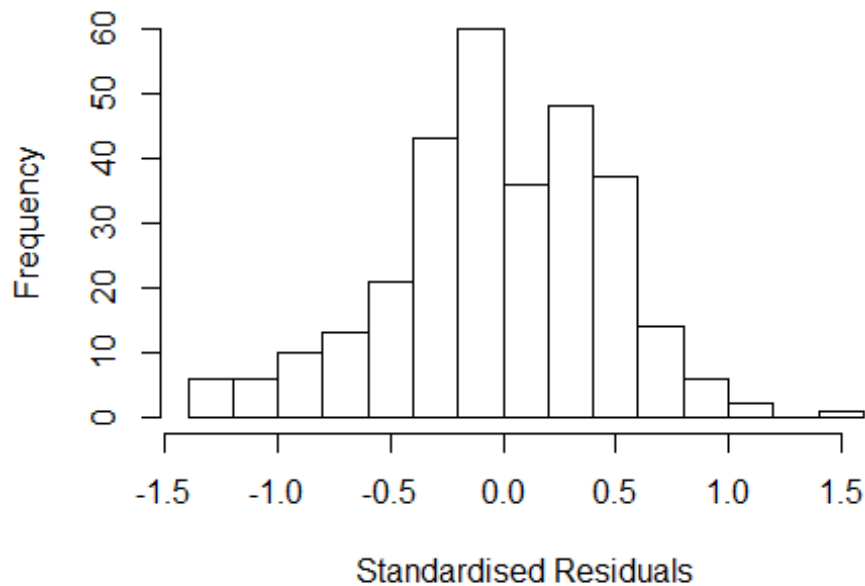
```

## Year2007          0.15484    0.31455    0.49    0.6229
## Year2008          0.13756    0.32046    0.43    0.6681
## Year2009          0.00376    0.35103    0.01    0.9915
## Year2010          0.20928    0.34489    0.61    0.5445
## Year2011          0.01160    0.32120    0.04    0.9712
## Year2012          0.20147    0.33421    0.60    0.5471
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.123, Adjusted R-squared:  0.071
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 277 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.340  0.882  0.953  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.312 1      1.145
## Year              1.312 16      1.009

```



## Residuals from last author



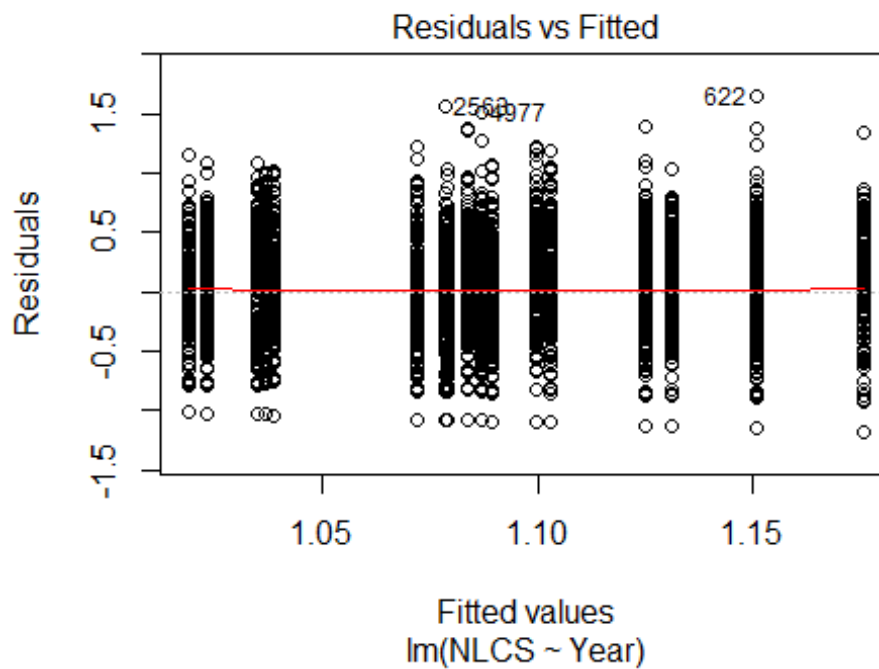
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3553 -0.2867 -0.0437 0.3363 1.4465
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9595 0.3063 3.13 0.0019 **
## LastAuthorFemale1 0.0213 0.0718 0.30 0.7668
## Year1997 0.3745 0.3728 1.00 0.3159
## Year1998 0.3229 0.4084 0.79 0.4299
## Year1999 0.6257 0.3237 1.93 0.0542 .
## Year2000 0.0123 0.3308 0.04 0.9703
## Year2001 0.0312 0.3236 0.10 0.9233
## Year2002 -0.1025 0.3381 -0.30 0.7620
## Year2003 0.4433 0.3210 1.38 0.1683
## Year2004 0.1608 0.3272 0.49 0.6235
## Year2005 0.3662 0.3191 1.15 0.2520
## Year2006 0.1512 0.3169 0.48 0.6336
```

```

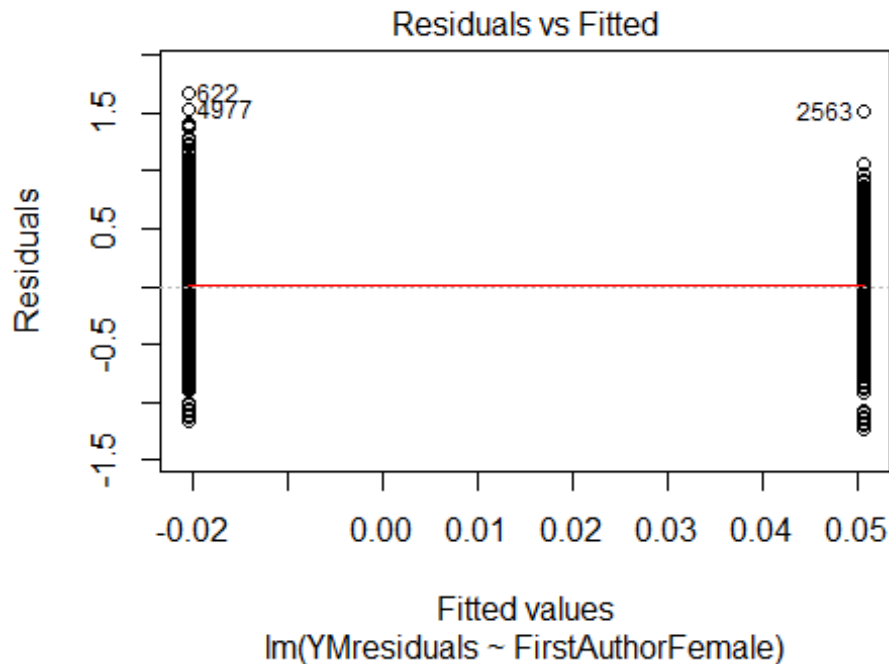
## Year2007          0.1808      0.3164      0.57      0.5682
## Year2008          0.1634      0.3220      0.51      0.6123
## Year2009          0.0402      0.3531      0.11      0.9093
## Year2010          0.2461      0.3439      0.72      0.4748
## Year2011          0.0381      0.3232      0.12      0.9064
## Year2012          0.2314      0.3415      0.68      0.4985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.115, Adjusted R-squared:  0.0625
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.341  0.885   0.955   0.907   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.30e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 303"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1602"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 450 499 352 399 329 313 330 286 332 299 385 338 313 353 355
## 2011 2012
## 397 369
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 272 289 179 233 175 119 226 203 230 204 250 243 222 254 263
## 2011 2012

```

```
## 299 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 247 252 166 207 154 109 195 177 198 182 222 221 199 234 232
## 2011 2012
## 273 233
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.06
```

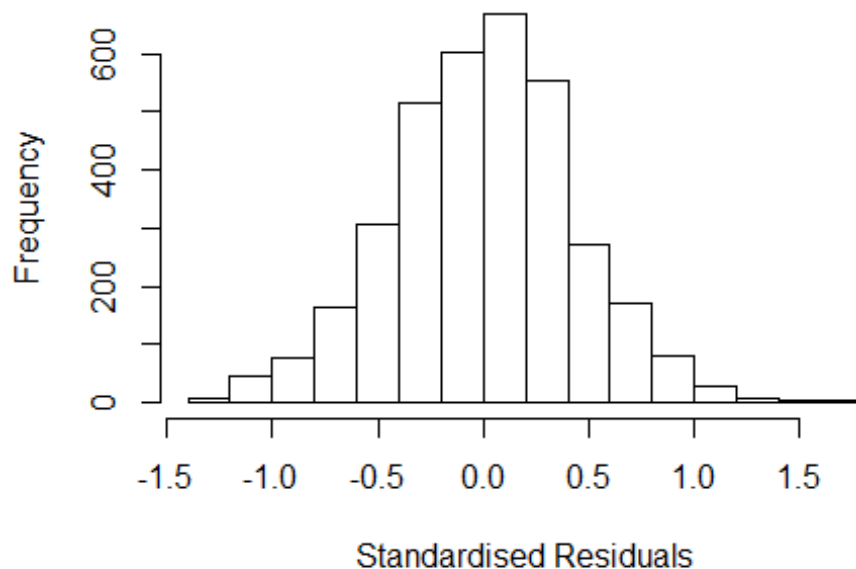


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 46, df = 1, p-value = 1e-11
```



```
## [1] "Female first author team size 2018 geometric mean: 3.92753541693997"
## [1] "Male first author team size 2018 geometric mean: 3.46795903533444"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3400, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.23536698851231"
## [1] "Male last author team size 2018 geometric mean: 3.70575920268993"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.050 1          1.025
## LastAuthorFemale  1.030 1          1.015
## UniqueAuthors    1.163 4          1.019
## Year              1.199 16         1.006
```

## Residuals from first and last author and team size



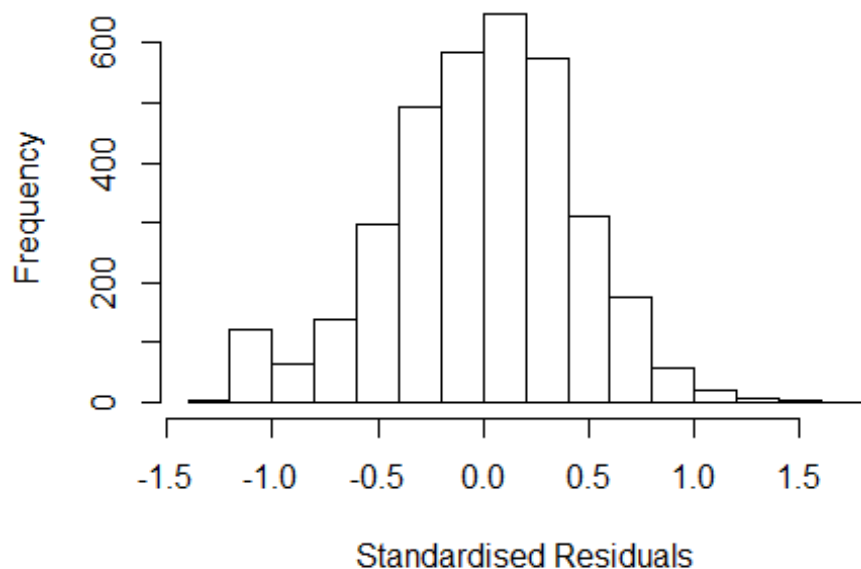
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2618 -0.2856 0.0117 0.2800 1.7155
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.80419 0.04949 16.25 < 2e-16 ***
## FirstAuthorFemale1 0.03714 0.01532 2.42 0.015 *
## LastAuthorFemale1 0.01499 0.02129 0.70 0.481
## UniqueAuthors2 0.28186 0.04369 6.45 1.3e-10 ***
## UniqueAuthors3 0.30603 0.04190 7.30 3.5e-13 ***
## UniqueAuthors4 0.35942 0.04291 8.38 < 2e-16 ***
## UniqueAuthors5 0.40928 0.04199 9.75 < 2e-16 ***
## Year1997 0.07494 0.04334 1.73 0.084 .
## Year1998 0.09815 0.04916 2.00 0.046 *
## Year1999 0.01473 0.04198 0.35 0.726
```

```

## Year2000      0.00927    0.05074    0.18    0.855
## Year2001     -0.03936    0.04767   -0.83    0.409
## Year2002     -0.03705    0.04213   -0.88    0.379
## Year2003     -0.08061    0.04515   -1.79    0.074 .
## Year2004     -0.03371    0.04666   -0.72    0.470
## Year2005     -0.10103    0.04310   -2.34    0.019 *
## Year2006     -0.10215    0.04183   -2.44    0.015 *
## Year2007     -0.04019    0.04248   -0.95    0.344
## Year2008      0.00431    0.04295    0.10    0.920
## Year2009     -0.09734    0.04290   -2.27    0.023 *
## Year2010     -0.05867    0.04125   -1.42    0.155
## Year2011     -0.10417    0.04130   -2.52    0.012 *
## Year2012     -0.10294    0.04304   -2.39    0.017 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.419
## Multiple R-squared:  0.0764, Adjusted R-squared:  0.0705
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 307 weights are ~= 1. The remaining 3194 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0561 0.8680 0.9500 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.86e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1 1.021
## LastAuthorFemale 1.025 1 1.012
## Year 1.062 16 1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2294 -0.2843 0.0186 0.2848 1.6436
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09820 0.03155 34.81 < 2e-16 ***
## FirstAuthorFemale1 0.06448 0.01561 4.13 3.7e-05 ***
## LastAuthorFemale1 0.01863 0.02162 0.86 0.389
## Year1997 0.05621 0.04276 1.31 0.189
## Year1998 0.06675 0.04857 1.37 0.169
## Year1999 0.00522 0.04214 0.12 0.901
## Year2000 0.00118 0.05305 0.02 0.982
## Year2001 -0.03524 0.04829 -0.73 0.466
## Year2002 -0.01590 0.04193 -0.38 0.705
## Year2003 -0.05018 0.04585 -1.09 0.274
## Year2004 -0.01435 0.04691 -0.31 0.760
## Year2005 -0.08699 0.04417 -1.97 0.049 *
```

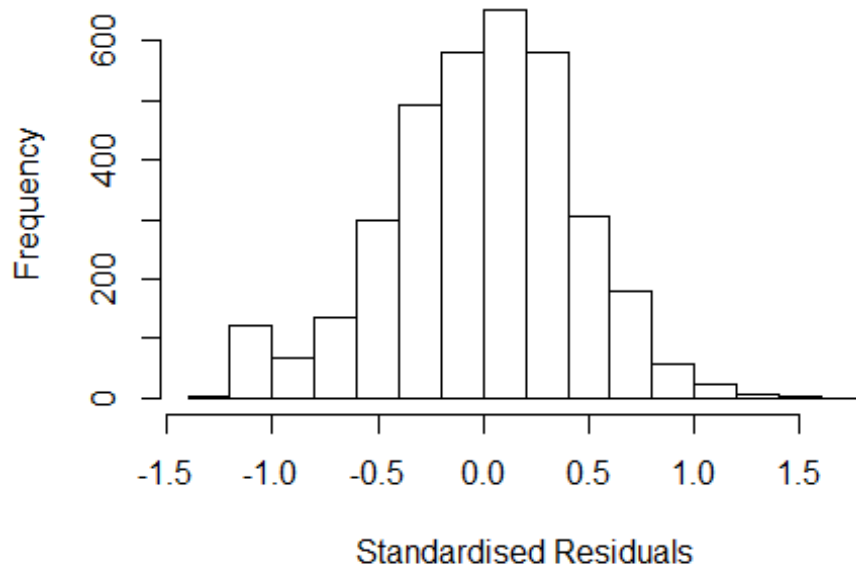
```

## Year2006      -0.07812    0.04241   -1.84    0.066 .
## Year2007      -0.00992    0.04257   -0.23    0.816
## Year2008       0.04002    0.04350    0.92    0.358
## Year2009      -0.07159    0.04475   -1.60    0.110
## Year2010      -0.02551    0.04197   -0.61    0.543
## Year2011      -0.06965    0.04163   -1.67    0.094 .
## Year2012      -0.05623    0.04352   -1.29    0.196
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared:  0.0141, Adjusted R-squared:  0.00897
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 254 weights are ~= 1. The remaining 3247 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  0.869  0.952  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.86e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1      1.019
## Year              1.038 16      1.001

```



## Residuals from first author



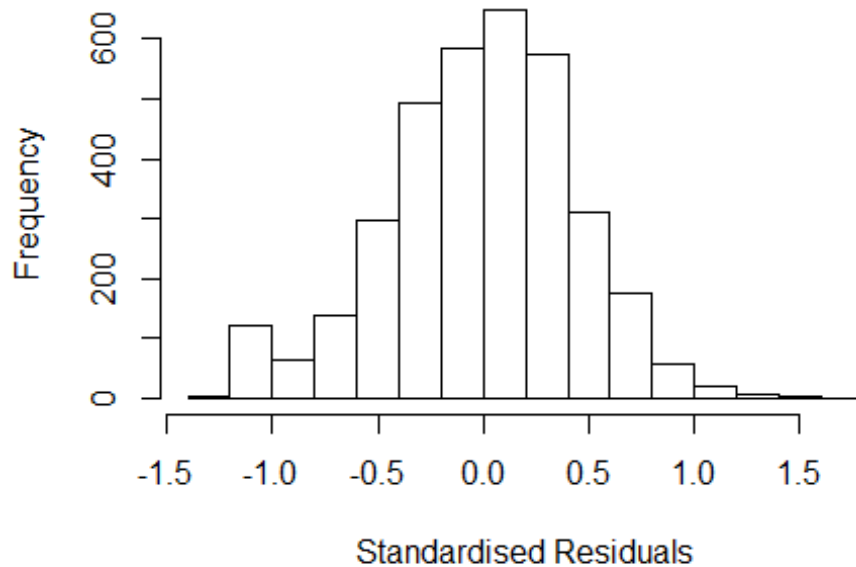
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2333 -0.2854 0.0179 0.2863 1.6415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.100703 0.031303 35.16 < 2e-16 ***
## FirstAuthorFemale1 0.065627 0.015594 4.21 2.6e-05 ***
## Year1997 0.055786 0.042745 1.31 0.192
## Year1998 0.066951 0.048545 1.38 0.168
## Year1999 0.005029 0.042155 0.12 0.905
## Year2000 0.000529 0.052995 0.01 0.992
## Year2001 -0.037129 0.048082 -0.77 0.440
## Year2002 -0.016196 0.041916 -0.39 0.699
## Year2003 -0.050278 0.045915 -1.10 0.274
## Year2004 -0.015071 0.046890 -0.32 0.748
## Year2005 -0.087588 0.044099 -1.99 0.047 *
## Year2006 -0.078006 0.042408 -1.84 0.066 .
```

```

## Year2007          -0.009950    0.042565   -0.23    0.815
## Year2008          0.040517    0.043454    0.93    0.351
## Year2009         -0.071186    0.044769   -1.59    0.112
## Year2010         -0.025287    0.041965   -0.60    0.547
## Year2011         -0.069306    0.041633   -1.66    0.096 .
## Year2012         -0.056377    0.043509   -1.30    0.195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared:  0.0139, Adjusted R-squared:  0.00904
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 3246 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.105  0.870  0.952  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.86e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1          1.011
## Year            1.021 16          1.001

```

## Residuals from last author



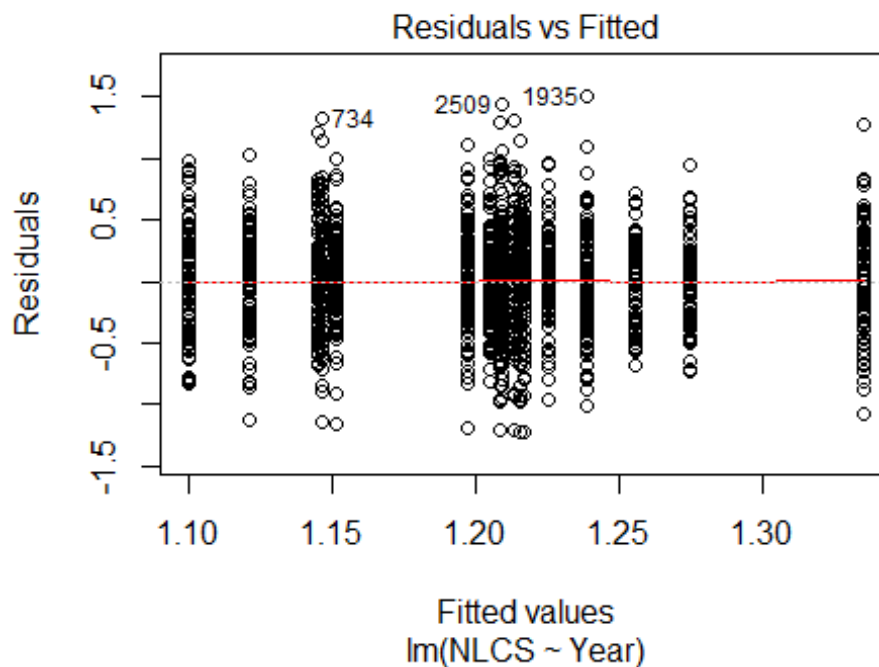
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1863 -0.2890  0.0197  0.2862  1.6330
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10992    0.03132   35.44  <2e-16 ***
## LastAuthorFemale1 0.02539    0.02167    1.17   0.241
## Year1997        0.05508    0.04282    1.29   0.198
## Year1998        0.06909    0.04912    1.41   0.160
## Year1999        0.00728    0.04210    0.17   0.863
## Year2000        0.00684    0.05304    0.13   0.897
## Year2001       -0.02953    0.04844   -0.61   0.542
## Year2002       -0.00832    0.04198   -0.20   0.843
## Year2003       -0.03772    0.04531   -0.83   0.405
## Year2004       -0.00537    0.04713   -0.11   0.909
## Year2005       -0.07420    0.04394   -1.69   0.091 .
## Year2006       -0.06681    0.04214   -1.59   0.113
```

```

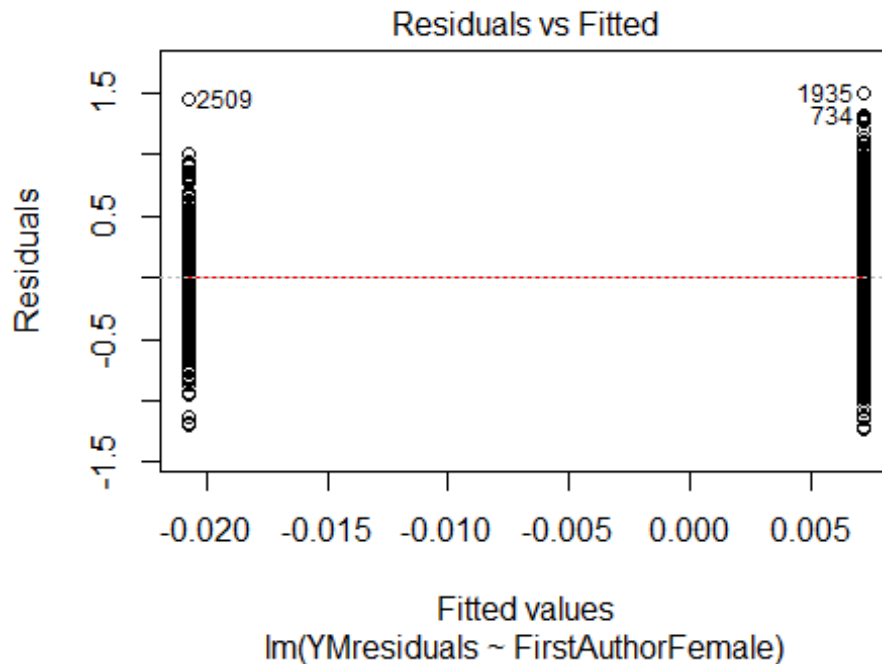
## Year2007      0.00152    0.04249    0.04    0.972
## Year2008      0.05098    0.04349    1.17    0.241
## Year2009     -0.06264    0.04466   -1.40    0.161
## Year2010     -0.01605    0.04183   -0.38    0.701
## Year2011     -0.05917    0.04143   -1.43    0.153
## Year2012     -0.04675    0.04336   -1.08    0.281
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.427
## Multiple R-squared:  0.00955,    Adjusted R-squared:  0.00471
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 262 weights are ~= 1. The remaining 3239 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.870  0.952  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.86e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3501"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1603"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 165 174 181 177 214 176 229 151 199 187 190 199 192 200 210
## 2011 2012
## 206 210
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 77 92 83 86 83 65 127 82 129 109 111 104 106 115 119
## 2011 2012

```

```
## 127 133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 83 75 74 73 57 106 73 113 99 107 96 97 102 104
## 2011 2012
## 119 119
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 34, df = 16, p-value = 0.005
```



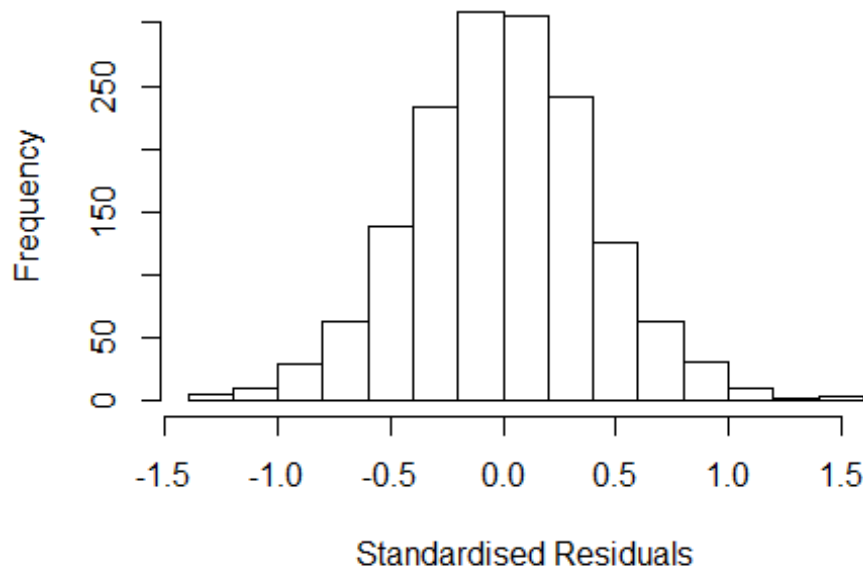
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 4.54148036446457"
## [1] "Male first author team size 2018 geometric mean: 3.44151028585921"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 650, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.10272223628024"
## [1] "Male last author team size 2018 geometric mean: 3.73741206331647"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 350, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.034	1	1.017
LastAuthorFemale	1.057	1	1.028
UniqueAuthors	1.327	4	1.036
Year	1.371	16	1.010

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27841 -0.26175 -0.00165  0.25704  1.49129
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1477     0.0844   13.60  <2e-16 ***
## FirstAuthorFemale1 -0.0249     0.0223   -1.12   0.2647
## LastAuthorFemale1 -0.0223     0.0367   -0.61   0.5435
## UniqueAuthors2     0.1307     0.0716    1.82   0.0683 .
## UniqueAuthors3     0.0970     0.0708    1.37   0.1709
## UniqueAuthors4     0.1502     0.0721    2.08   0.0373 *
## UniqueAuthors5     0.1481     0.0706    2.10   0.0361 *
## Year1997          -0.0326     0.0714   -0.46   0.6482
## Year1998           0.1102     0.0740    1.49   0.1363
## Year1999          -0.1037     0.0783   -1.32   0.1857
```

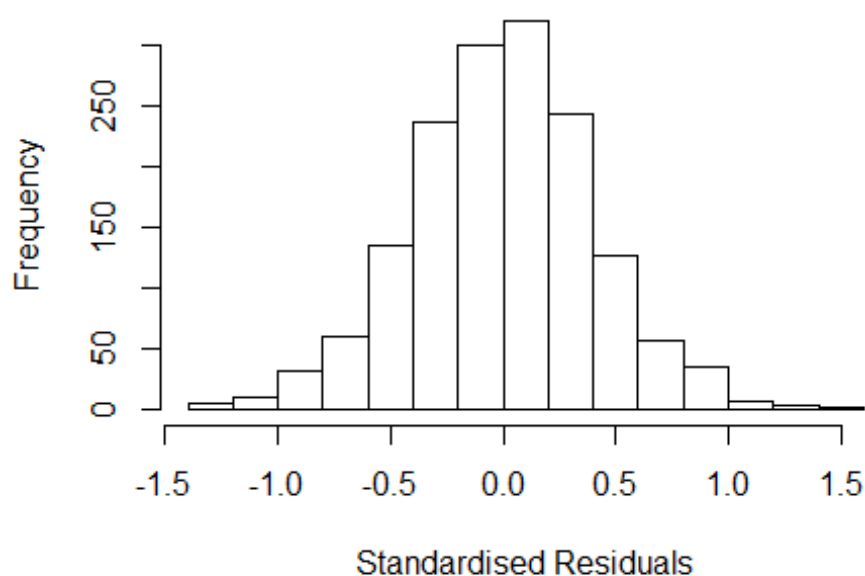
```

## Year2000          0.0147      0.0709      0.21      0.8355
## Year2001         -0.0198      0.0715     -0.28      0.7823
## Year2002         -0.0579      0.0653     -0.89      0.3748
## Year2003         -0.1415      0.0722     -1.96      0.0502 .
## Year2004         -0.0307      0.0656     -0.47      0.6396
## Year2005         -0.0678      0.0698     -0.97      0.3317
## Year2006         -0.0379      0.0694     -0.55      0.5852
## Year2007         -0.0748      0.0696     -1.07      0.2829
## Year2008         -0.0317      0.0653     -0.49      0.6273
## Year2009         -0.1329      0.0645     -2.06      0.0393 *
## Year2010         -0.0567      0.0644     -0.88      0.3784
## Year2011         -0.1478      0.0670     -2.21      0.0275 *
## Year2012         -0.1782      0.0653     -2.73      0.0064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0341, Adjusted R-squared:  0.0204
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1409 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.101  0.867  0.948  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1      1.015
## LastAuthorFemale  1.034 1      1.017
## Year              1.060 16      1.002

```



## Residuals from first and last author



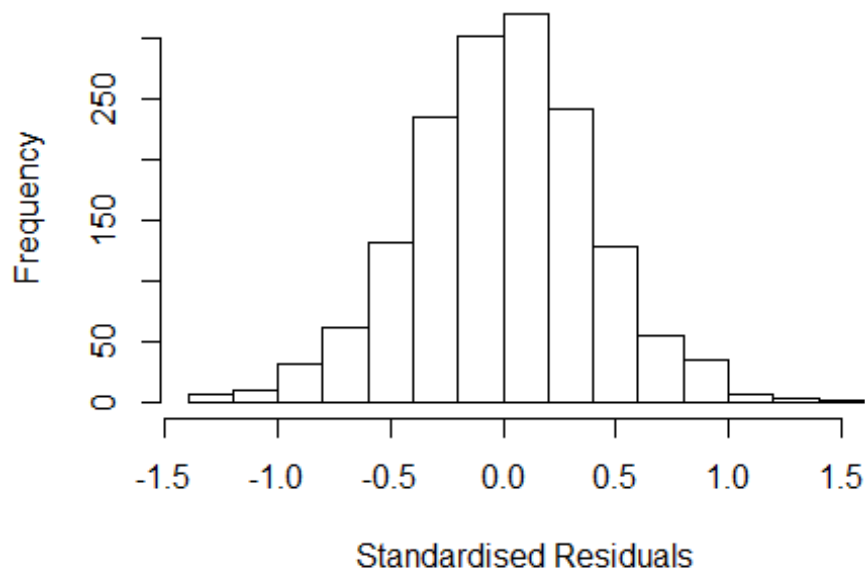
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26662 -0.27075  0.00268  0.26109  1.49541
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2666    0.0533   23.76  <2e-16 ***
## FirstAuthorFemale1 -0.0211    0.0225   -0.93   0.350
## LastAuthorFemale1  -0.0282    0.0366   -0.77   0.442
## Year1997          -0.0352    0.0711   -0.50   0.620
## Year1998           0.1076    0.0745    1.44   0.149
## Year1999          -0.0998    0.0781   -1.28   0.202
## Year2000           0.0226    0.0706    0.32   0.749
## Year2001          -0.0128    0.0704   -0.18   0.856
## Year2002          -0.0538    0.0653   -0.82   0.410
## Year2003          -0.1303    0.0725   -1.80   0.072 .
## Year2004          -0.0230    0.0651   -0.35   0.724
## Year2005          -0.0508    0.0688   -0.74   0.461
```

```

## Year2006          -0.0303      0.0689   -0.44    0.660
## Year2007          -0.0730      0.0682   -1.07    0.285
## Year2008          -0.0216      0.0653   -0.33    0.741
## Year2009          -0.1200      0.0636   -1.89    0.059 .
## Year2010          -0.0489      0.0640   -0.76    0.445
## Year2011          -0.1361      0.0661   -2.06    0.039 *
## Year2012          -0.1656      0.0644   -2.57    0.010 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0266, Adjusted R-squared:  0.0153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1416 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.105  0.870  0.948  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.03 1      1.015
## Year              1.03 16      1.001

```

## Residuals from first author



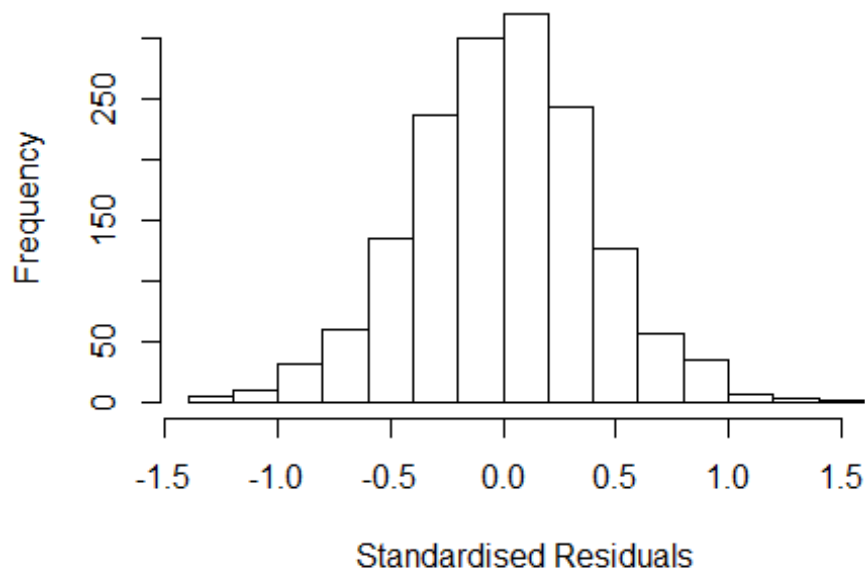
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2659 -0.2711  0.0034  0.2611  1.4969
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2659     0.0534   23.72  <2e-16 ***
## FirstAuthorFemale1 -0.0227     0.0225   -1.01   0.3138
## Year1997         -0.0379     0.0708   -0.54   0.5921
## Year1998          0.1058     0.0744    1.42   0.1556
## Year1999         -0.1007     0.0780   -1.29   0.1971
## Year2000          0.0213     0.0706    0.30   0.7629
## Year2001         -0.0128     0.0705   -0.18   0.8564
## Year2002         -0.0553     0.0652   -0.85   0.3969
## Year2003         -0.1323     0.0723   -1.83   0.0676 .
## Year2004         -0.0238     0.0651   -0.37   0.7142
## Year2005         -0.0523     0.0687   -0.76   0.4464
## Year2006         -0.0303     0.0689   -0.44   0.6600
```

```

## Year2007          -0.0751      0.0681   -1.10    0.2705
## Year2008          -0.0242      0.0653   -0.37    0.7113
## Year2009          -0.1221      0.0635   -1.92    0.0547 .
## Year2010          -0.0502      0.0639   -0.79    0.4319
## Year2011          -0.1368      0.0660   -2.07    0.0384 *
## Year2012          -0.1664      0.0643   -2.59    0.0098 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0263, Adjusted R-squared:  0.0156
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 150 weights are ~= 1. The remaining 1417 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  0.870  0.948  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.031 1          1.015
## Year            1.031 16          1.001

```

## Residuals from last author



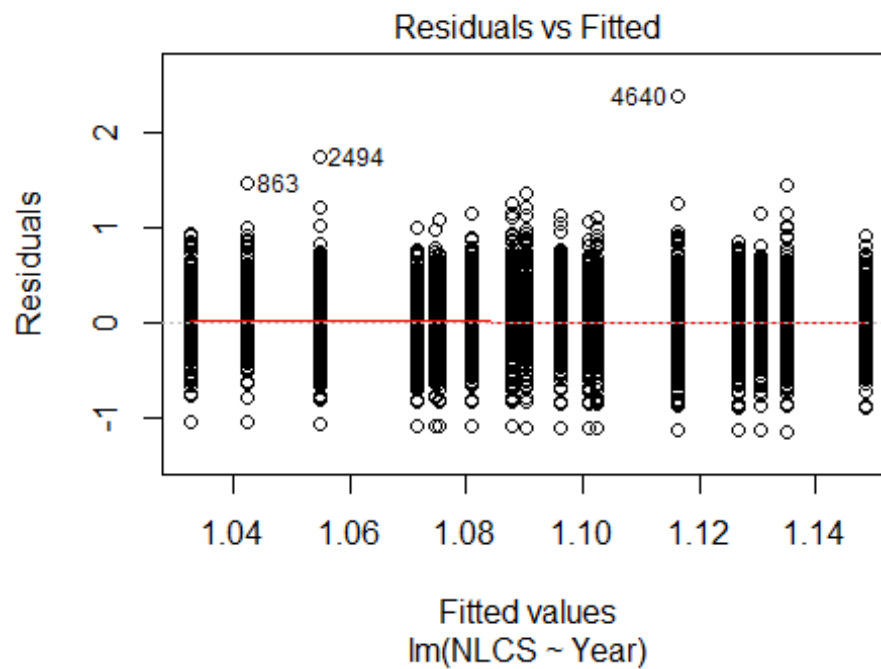
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26427 -0.26622  0.00594  0.25882  1.50097
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2643     0.0530   23.87  <2e-16 ***
## LastAuthorFemale1 -0.0312     0.0366   -0.85    0.395
## Year1997         -0.0348     0.0708   -0.49    0.623
## Year1998          0.1057     0.0744    1.42    0.156
## Year1999         -0.1022     0.0783   -1.30    0.192
## Year2000          0.0196     0.0705    0.28    0.781
## Year2001         -0.0147     0.0703   -0.21    0.834
## Year2002         -0.0564     0.0651   -0.87    0.386
## Year2003         -0.1341     0.0723   -1.85    0.064 .
## Year2004         -0.0262     0.0649   -0.40    0.686
## Year2005         -0.0550     0.0685   -0.80    0.422
## Year2006         -0.0348     0.0689   -0.50    0.614
```

```

## Year2007          -0.0745      0.0680    -1.09      0.274
## Year2008          -0.0247      0.0650    -0.38      0.704
## Year2009          -0.1248      0.0634    -1.97      0.049 *
## Year2010          -0.0526      0.0637    -0.82      0.410
## Year2011          -0.1402      0.0658    -2.13      0.033 *
## Year2012          -0.1701      0.0640    -2.66      0.008 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.0261, Adjusted R-squared:  0.0154
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~ = 1. The remaining 1416 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.869  0.948  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1567"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1604"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  501  412  415  387  377  334  313  411  408  342  400  350  329  378  383
## 2011 2012
##  339  306
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  298  271  295  270  254  207  237  339  344  289  330  277  252  300  306
## 2011 2012

```

```
## 267 235
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 256 222 258 248 221 180 212 306 306 253 302 252 220 281 281
## 2011 2012
## 243 206
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 69, df = 16, p-value = 1e-08
```

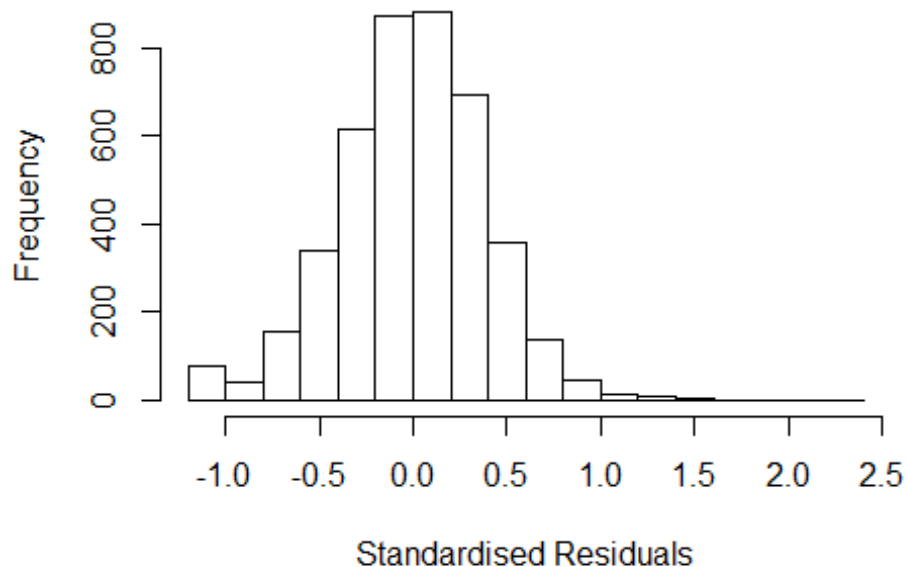


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```





## Residuals from first and last author and team size



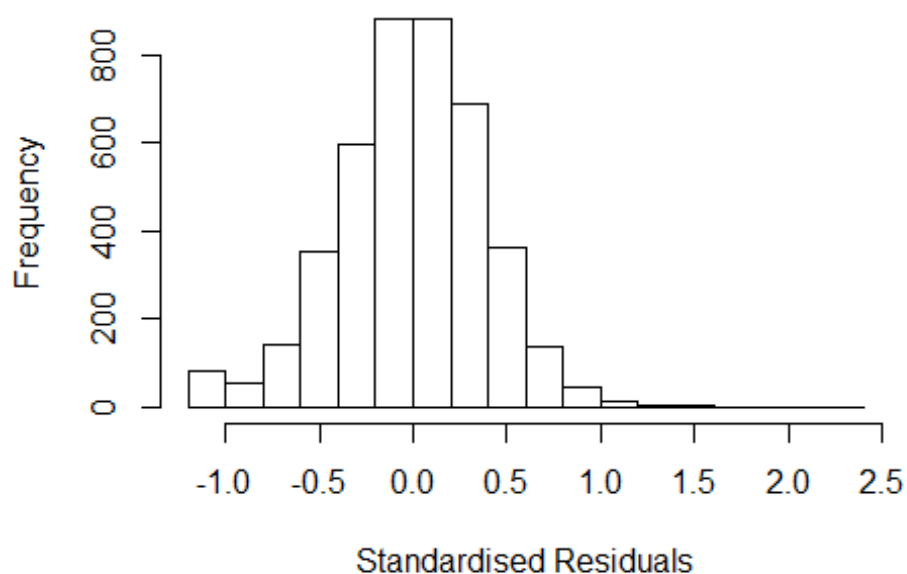
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18760 -0.24355 0.00534 0.25601 2.32982
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.019790 0.062186 16.40 <2e-16 ***
## FirstAuthorFemale1 0.016023 0.014525 1.10 0.270
## LastAuthorFemale1 -0.026795 0.017432 -1.54 0.124
## UniqueAuthors2 0.032289 0.057508 0.56 0.575
## UniqueAuthors3 0.040387 0.056330 0.72 0.473
## UniqueAuthors4 0.087066 0.056542 1.54 0.124
## UniqueAuthors5 0.126692 0.055977 2.26 0.024 *
## Year1997 -0.030021 0.036278 -0.83 0.408
## Year1998 -0.017698 0.033873 -0.52 0.601
## Year1999 0.011853 0.036980 0.32 0.749
```

```

## Year2000      -0.033851    0.034694   -0.98    0.329
## Year2001      -0.025376    0.037969   -0.67    0.504
## Year2002      -0.009488    0.036412   -0.26    0.794
## Year2003       0.032159    0.032822    0.98    0.327
## Year2004       0.041122    0.033791    1.22    0.224
## Year2005       0.020700    0.035934    0.58    0.565
## Year2006      -0.000827    0.032568   -0.03    0.980
## Year2007       0.051346    0.034318    1.50    0.135
## Year2008      -0.020093    0.037176   -0.54    0.589
## Year2009      -0.004687    0.037877   -0.12    0.902
## Year2010       0.037428    0.036592    1.02    0.306
## Year2011       0.005299    0.037814    0.14    0.889
## Year2012      -0.003094    0.039564   -0.08    0.938
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.367
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.014
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1168,2334) are outliers with |weight| = 0 ( < 2.4e-05);
## 368 weights are ~= 1. The remaining 3877 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0571 0.8690 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.095 1 1.046
## LastAuthorFemale 1.017 1 1.008
## Year 1.105 16 1.003

```

## Residuals from first and last author



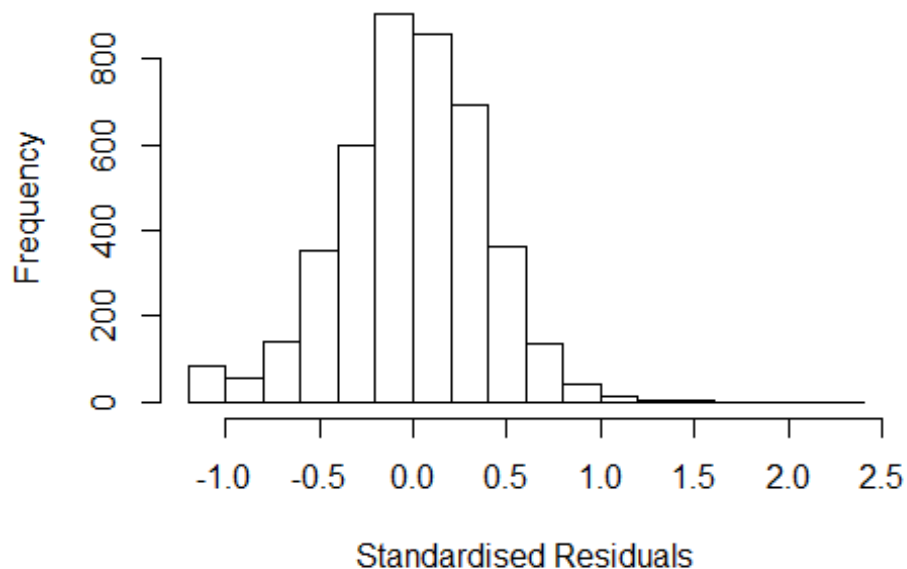
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1633 -0.2453 0.0023 0.2526 2.3790
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09e+00 2.57e-02 42.19 <2e-16 ***
## FirstAuthorFemale1 1.90e-02 1.46e-02 1.30 0.192
## LastAuthorFemale1 -2.97e-02 1.74e-02 -1.71 0.088 .
## Year1997 -2.96e-02 3.64e-02 -0.81 0.416
## Year1998 -5.76e-03 3.37e-02 -0.17 0.864
## Year1999 1.86e-02 3.64e-02 0.51 0.608
## Year2000 -2.61e-02 3.45e-02 -0.76 0.450
## Year2001 -2.19e-02 3.76e-02 -0.58 0.560
## Year2002 -5.58e-05 3.63e-02 0.00 0.999
## Year2003 4.83e-02 3.27e-02 1.48 0.140
## Year2004 5.70e-02 3.35e-02 1.70 0.089 .
## Year2005 3.29e-02 3.61e-02 0.91 0.362
```

```

## Year2006          1.80e-02   3.21e-02   0.56   0.576
## Year2007          7.01e-02   3.38e-02   2.08   0.038 *
## Year2008          3.22e-03   3.70e-02   0.09   0.931
## Year2009          1.78e-02   3.79e-02   0.47   0.639
## Year2010          5.92e-02   3.61e-02   1.64   0.101
## Year2011          2.98e-02   3.75e-02   0.79   0.427
## Year2012          2.01e-02   3.91e-02   0.51   0.607
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.369
## Multiple R-squared:  0.00728,    Adjusted R-squared:  0.00305
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1168,2334) are outliers with |weight| = 0 ( < 2.4e-05);
## 358 weights are ~ = 1. The remaining 3887 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0808 0.8710 0.9500 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1          1.044
## Year              1.091 16          1.003

```

## Residuals from first author



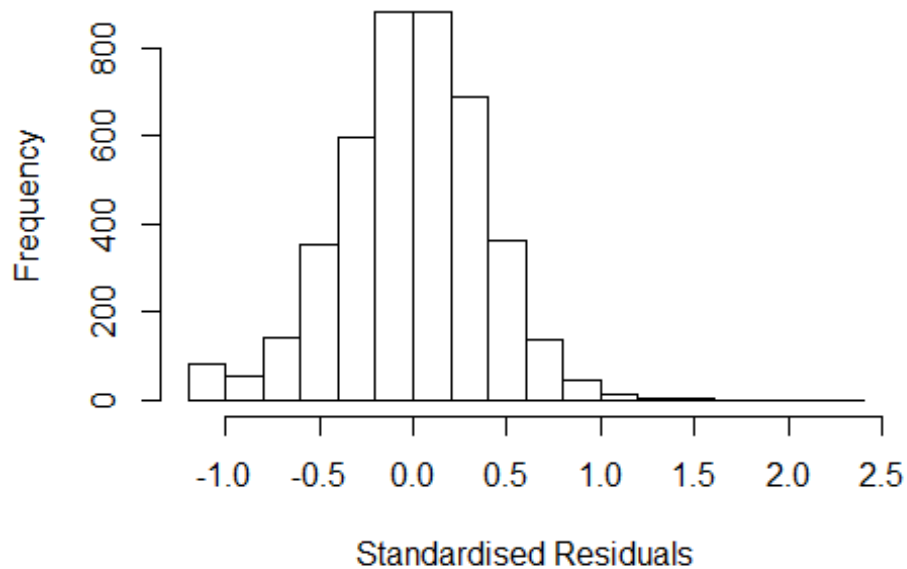
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.15803 -0.24592 -0.00234 0.25377 2.38383
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08275 0.02562 42.26 <2e-16 ***
## FirstAuthorFemale1 0.01737 0.01458 1.19 0.234
## Year1997 -0.03001 0.03640 -0.82 0.410
## Year1998 -0.00639 0.03366 -0.19 0.849
## Year1999 0.01655 0.03634 0.46 0.649
## Year2000 -0.02862 0.03443 -0.83 0.406
## Year2001 -0.02286 0.03756 -0.61 0.543
## Year2002 -0.00105 0.03626 -0.03 0.977
## Year2003 0.04718 0.03274 1.44 0.150
## Year2004 0.05580 0.03344 1.67 0.095 .
## Year2005 0.03043 0.03599 0.85 0.398
## Year2006 0.01708 0.03205 0.53 0.594
```

```

## Year2007          0.06854    0.03365    2.04    0.042 *
## Year2008          0.00193    0.03683    0.05    0.958
## Year2009          0.01673    0.03789    0.44    0.659
## Year2010          0.05791    0.03606    1.61    0.108
## Year2011          0.02773    0.03741    0.74    0.459
## Year2012          0.01859    0.03906    0.48    0.634
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.369
## Multiple R-squared:  0.00656,    Adjusted R-squared:  0.00256
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1168,2334) are outliers with |weight| = 0 ( < 2.4e-05);
## 370 weights are ~= 1. The remaining 3875 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0795 0.8710 0.9500 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.015 1          1.007
## Year          1.015 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14967 -0.24667 0.00187 0.25352 2.37455
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09e+00 2.57e-02 42.37 <2e-16 ***
## LastAuthorFemale1 -2.81e-02 1.74e-02 -1.62 0.106
## Year1997 -2.88e-02 3.64e-02 -0.79 0.428
## Year1998 -5.76e-03 3.37e-02 -0.17 0.864
## Year1999 1.95e-02 3.64e-02 0.54 0.592
## Year2000 -2.75e-02 3.45e-02 -0.80 0.425
## Year2001 -2.24e-02 3.76e-02 -0.60 0.551
## Year2002 -6.78e-05 3.63e-02 0.00 0.999
## Year2003 4.97e-02 3.28e-02 1.52 0.129
## Year2004 5.88e-02 3.34e-02 1.76 0.079 .
## Year2005 3.48e-02 3.61e-02 0.97 0.334
## Year2006 1.91e-02 3.21e-02 0.60 0.551
```

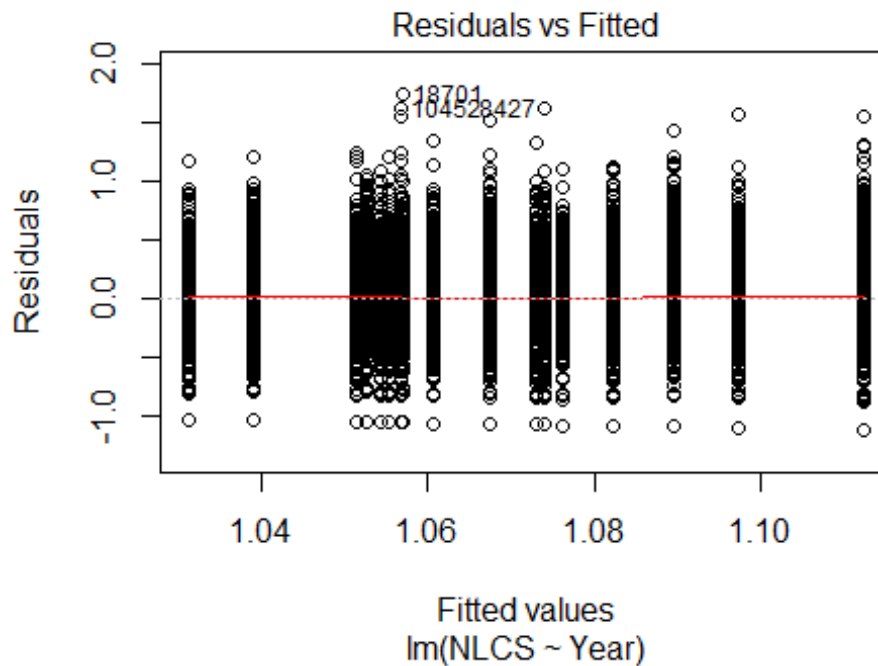
```

## Year2007          7.26e-02   3.37e-02   2.15   0.031 *
## Year2008          5.20e-03   3.69e-02   0.14   0.888
## Year2009          2.06e-02   3.78e-02   0.55   0.586
## Year2010          6.21e-02   3.60e-02   1.72   0.085 .
## Year2011          3.37e-02   3.72e-02   0.91   0.364
## Year2012          2.46e-02   3.88e-02   0.63   0.527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.369
## Multiple R-squared:  0.00689,    Adjusted R-squared:  0.0029
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1168,2334) are outliers with |weight| = 0 ( < 2.4e-05);
## 371 weights are ~1. The remaining 3874 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0827 0.8710 0.9490 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4247"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1605"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1072 1058 989 954 1026 912 871 944 901 838 900 820 844 887 907
## 2011 2012
## 872 772
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 721 725 715 647 623 516 645 742 678 650 694 645 691 697 724

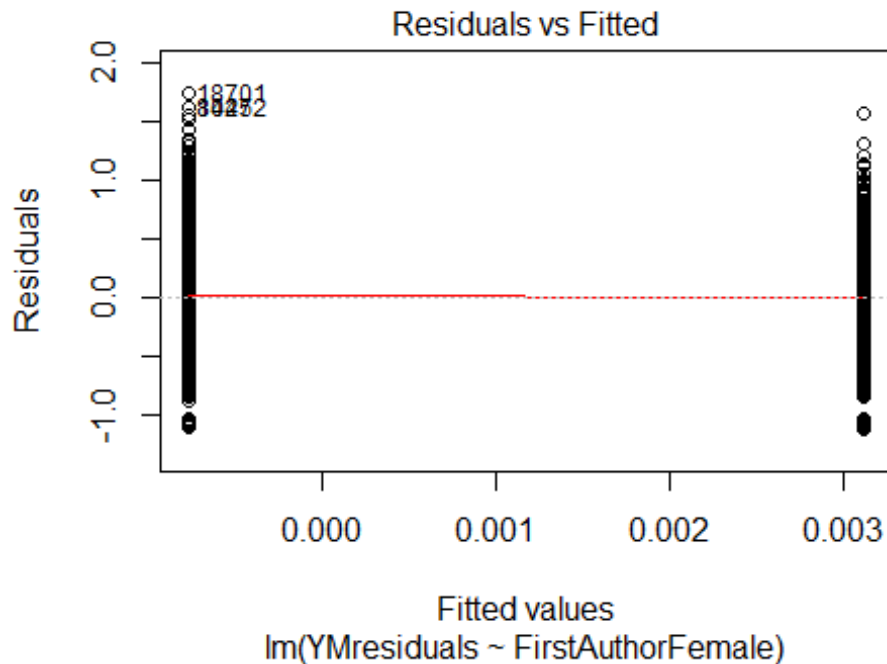
```



```
## 2011 2012
## 684 576
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 617 628 630 581 556 457 594 670 618 595 626 591 617 637 648
## 2011 2012
## 619 519
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

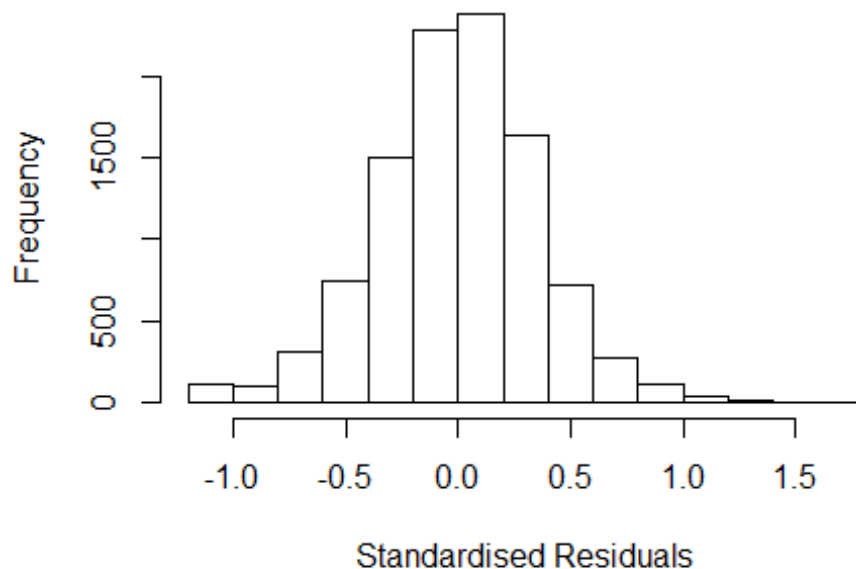


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 4.60411236304323"
## [1] "Male first author team size 2018 geometric mean: 4.32642748947738"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.9630425901562"
## [1] "Male last author team size 2018 geometric mean: 4.34707217940086"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5700, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.034 1          1.017
## LastAuthorFemale  1.012 1          1.006
## UniqueAuthors    1.081 4          1.010
## Year              1.121 16         1.004
```

## Residuals from first and last author and team size



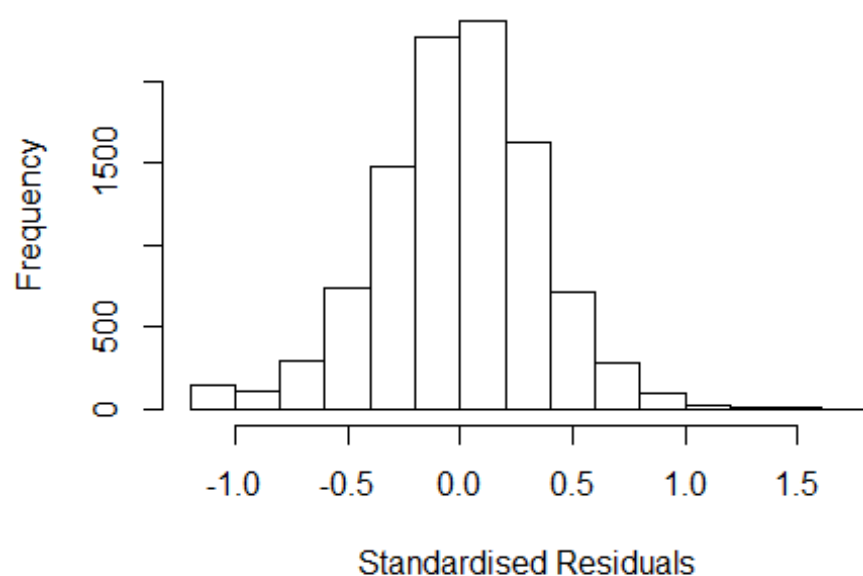
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18070 -0.22369 0.00396 0.21995 1.75042
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04426 0.03278 31.86 < 2e-16 ***
## FirstAuthorFemale1 -0.00230 0.00852 -0.27 0.78743
## LastAuthorFemale1 -0.03101 0.01067 -2.91 0.00368 **
## UniqueAuthors2 0.03779 0.02957 1.28 0.20126
## UniqueAuthors3 0.06462 0.02904 2.23 0.02610 *
## UniqueAuthors4 0.09870 0.02924 3.38 0.00074 ***
## UniqueAuthors5 0.13644 0.02873 4.75 2.1e-06 ***
## Year1997 -0.02401 0.02331 -1.03 0.30296
## Year1998 -0.01806 0.02271 -0.80 0.42642
## Year1999 -0.04114 0.02212 -1.86 0.06289 .
```

```

## Year2000      -0.07090      0.02216      -3.20      0.00138 **
## Year2001      -0.05460      0.02276      -2.40      0.01645 *
## Year2002      -0.06079      0.02187      -2.78      0.00545 **
## Year2003      -0.04890      0.02099      -2.33      0.01983 *
## Year2004      -0.06679      0.02061      -3.24      0.00120 **
## Year2005      -0.08527      0.02154      -3.96      7.6e-05 ***
## Year2006      -0.06466      0.02095      -3.09      0.00203 **
## Year2007      -0.07005      0.02207      -3.17      0.00151 **
## Year2008      -0.04605      0.02137      -2.16      0.03117 *
## Year2009      -0.09782      0.02227      -4.39      1.1e-05 ***
## Year2010      -0.06146      0.02234      -2.75      0.00596 **
## Year2011      -0.07572      0.02259      -3.35      0.00081 ***
## Year2012      -0.06630      0.02394      -2.77      0.00563 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0181, Adjusted R-squared:  0.016
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 6 observations c(578,1855,4245,5739,8746,10037)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 869 weights are ~ = 1. The remaining 9328 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0334 0.8670 0.9500 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.80e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1 1.015
## LastAuthorFemale 1.008 1 1.004
## Year 1.039 16 1.001

```

## Residuals from first and last author



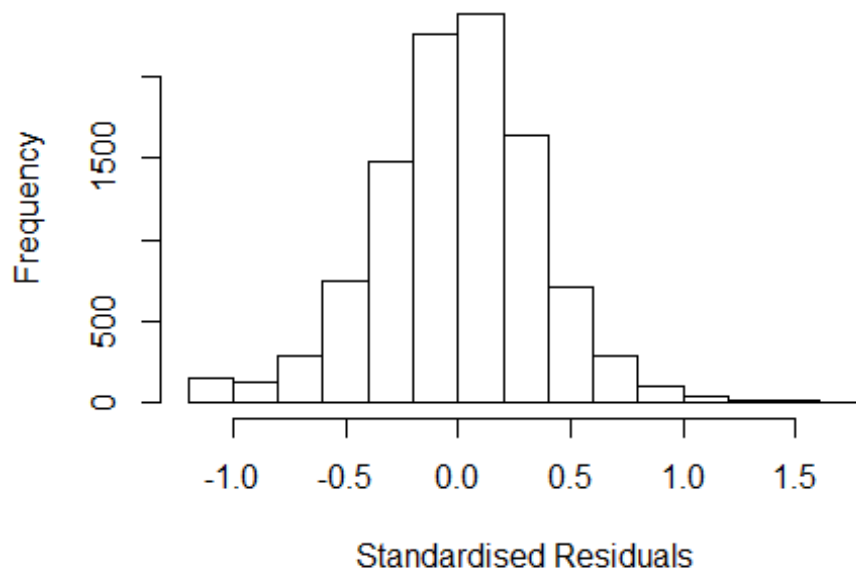
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12069 -0.22283  0.00288  0.22317  1.71602
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.120691    0.016722   67.02 < 2e-16 ***
## FirstAuthorFemale1 -0.000411    0.008599   -0.05  0.96187
## LastAuthorFemale1 -0.029244    0.010740   -2.72  0.00648 **
## Year1997        -0.022926    0.023217   -0.99  0.32342
## Year1998        -0.015172    0.022608   -0.67  0.50218
## Year1999        -0.034883    0.022143   -1.58  0.11521
## Year2000        -0.065574    0.022135   -2.96  0.00306 **
## Year2001        -0.047948    0.022789   -2.10  0.03540 *
## Year2002        -0.054717    0.021697   -2.52  0.01169 *
## Year2003        -0.034191    0.020894   -1.64  0.10179
## Year2004        -0.052139    0.020484   -2.55  0.01093 *
## Year2005        -0.070270    0.021372   -3.29  0.00101 **
```

```

## Year2006          -0.045873    0.020880    -2.20    0.02805 *
## Year2007          -0.050268    0.021943    -2.29    0.02199 *
## Year2008          -0.022930    0.021159    -1.08    0.27852
## Year2009          -0.075829    0.022228    -3.41    0.00065 ***
## Year2010          -0.042457    0.022271    -1.91    0.05663 .
## Year2011          -0.053699    0.022438    -2.39    0.01672 *
## Year2012          -0.043710    0.023962    -1.82    0.06816 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.00406,    Adjusted R-squared:  0.0023
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(1855,4245,10037)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 888 weights are ~ = 1. The remaining 9312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.000  0.866  0.951    0.893  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.80e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1          1.015
## Year              1.031 16          1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11756 -0.22263 0.00444 0.22314 1.71916
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11756 0.01666 67.08 < 2e-16 ***
## FirstAuthorFemale1 -0.00053 0.00861 -0.06 0.95093
## Year1997 -0.02206 0.02319 -0.95 0.34143
## Year1998 -0.01589 0.02261 -0.70 0.48214
## Year1999 -0.03544 0.02213 -1.60 0.10936
## Year2000 -0.06570 0.02212 -2.97 0.00298 **
## Year2001 -0.04814 0.02279 -2.11 0.03470 *
## Year2002 -0.05592 0.02167 -2.58 0.00987 **
## Year2003 -0.03484 0.02087 -1.67 0.09509 .
## Year2004 -0.05261 0.02048 -2.57 0.01021 *
## Year2005 -0.07100 0.02134 -3.33 0.00088 ***
## Year2006 -0.04609 0.02086 -2.21 0.02716 *
```

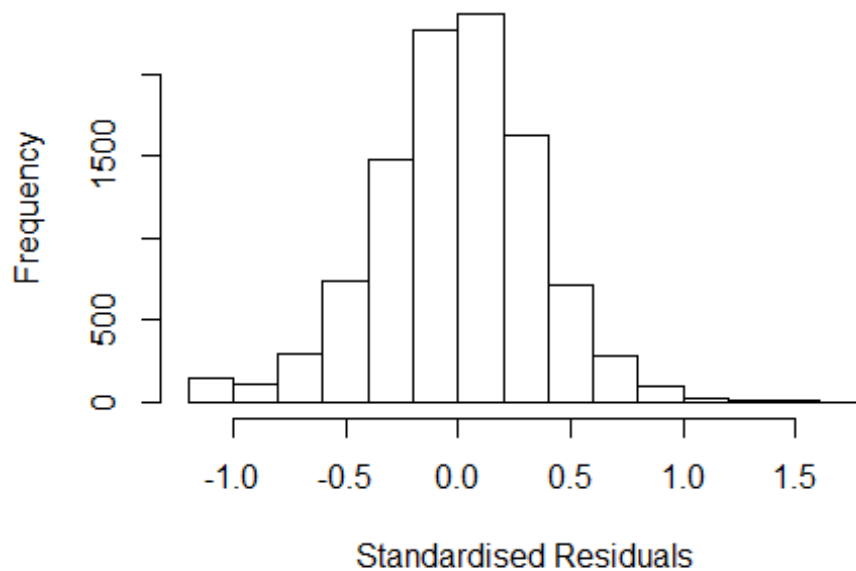
```

## Year2007      -0.05018    0.02192   -2.29  0.02208 *
## Year2008      -0.02387    0.02114   -1.13  0.25885
## Year2009      -0.07732    0.02222   -3.48  0.00050 ***
## Year2010      -0.04236    0.02225   -1.90  0.05687 .
## Year2011      -0.05382    0.02244   -2.40  0.01647 *
## Year2012      -0.04372    0.02396   -1.82  0.06806 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.00327,    Adjusted R-squared:  0.00161
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 4 observations c(1855,4245,5739,10037)
## are outliers with |weight| <= 4.6e-06 ( < 9.8e-06);
## 889 weights are ~ = 1. The remaining 9310 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8650 0.9500 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.80e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.008 1          1.004
## Year            1.008 16          1.000

```



## Residuals from last author



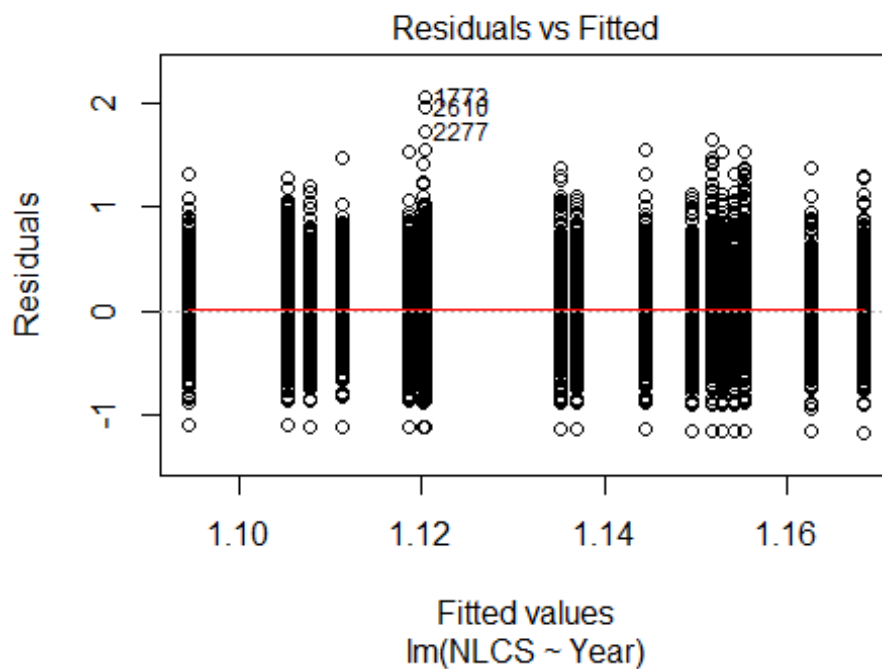
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12063 -0.22282  0.00296  0.22325  1.71614
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1206    0.0167   67.19 < 2e-16 ***
## LastAuthorFemale1 -0.0292    0.0107   -2.72  0.00648 **
## Year1997         -0.0229    0.0232   -0.99  0.32350
## Year1998         -0.0152    0.0226   -0.67  0.50228
## Year1999         -0.0349    0.0221   -1.58  0.11495
## Year2000         -0.0656    0.0221   -2.96  0.00304 **
## Year2001         -0.0479    0.0228   -2.10  0.03541 *
## Year2002         -0.0547    0.0217   -2.52  0.01167 *
## Year2003         -0.0342    0.0209   -1.64  0.10150
## Year2004         -0.0522    0.0205   -2.55  0.01091 *
## Year2005         -0.0703    0.0214   -3.29  0.00100 ***
## Year2006         -0.0459    0.0209   -2.20  0.02785 *
```

```

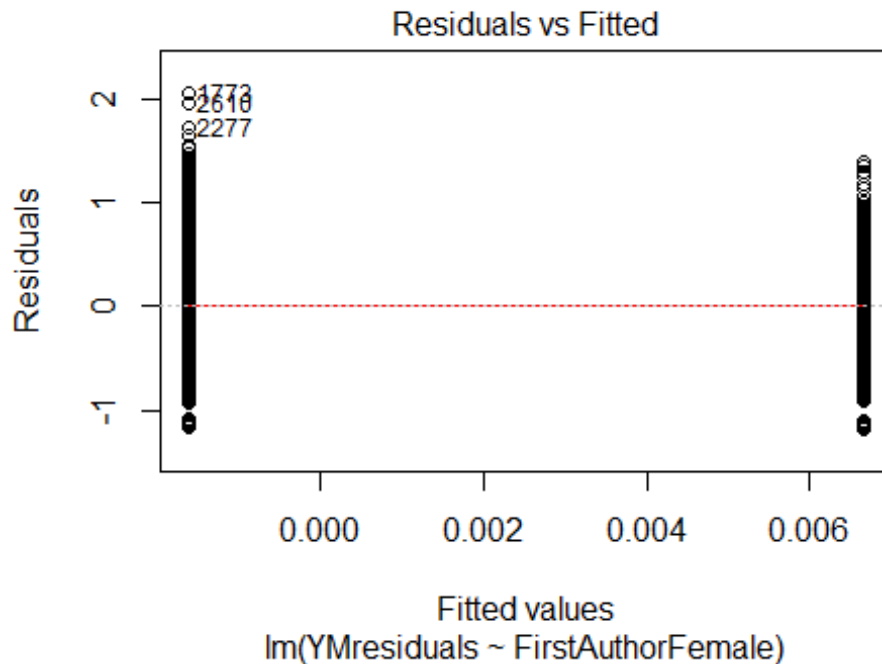
## Year2007          -0.0503      0.0219    -2.29   0.02188 *
## Year2008          -0.0230      0.0211    -1.09   0.27671
## Year2009          -0.0759      0.0222    -3.41   0.00064 ***
## Year2010          -0.0425      0.0222    -1.91   0.05592 .
## Year2011          -0.0538      0.0224    -2.40   0.01624 *
## Year2012          -0.0438      0.0239    -1.83   0.06708 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.332
## Multiple R-squared:  0.00406,    Adjusted R-squared:  0.0024
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(1855,4245,10037)
## are outliers with |weight| = 0 ( < 9.8e-06);
## 892 weights are ~= 1. The remaining 9308 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  0.866   0.951   0.893   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.80e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 10203"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1606"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1159 1111 1058 1055 1064  947  968 1072  979  944  988  940  928  951  936
## 2011 2012
##   964   809
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 646 617 585 606 531 444 579 745 672 642 658 606 619 671 672
## 2011 2012
## 676 584
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 573 542 528 549 481 401 529 685 607 577 600 550 549 608 592
## 2011 2012
## 604 513
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 94, df = 16, p-value = 4e-13
```

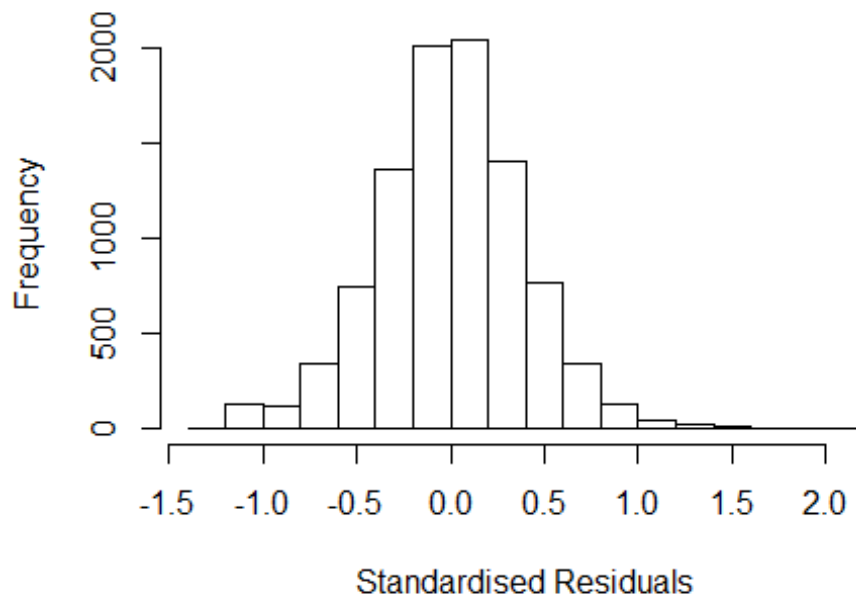


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 7e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.03025740157609"
## [1] "Male first author team size 2018 geometric mean: 3.39366475529688"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.09670951017553"
## [1] "Male last author team size 2018 geometric mean: 3.45959084790632"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1      1.014
## LastAuthorFemale  1.016 1      1.008
## UniqueAuthors     1.056 4      1.007
## Year              1.075 16      1.002
```

## Residuals from first and last author and team size



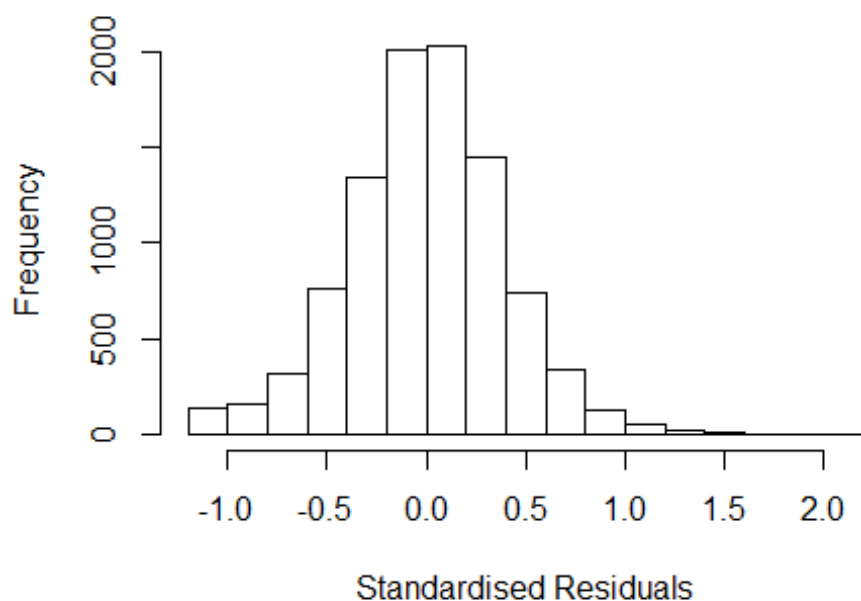
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21663 -0.23831 0.00157 0.24098 2.04122
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04379 0.02649 39.40 < 2e-16 ***
## FirstAuthorFemale1 0.01228 0.00957 1.28 0.19970
## LastAuthorFemale1 -0.04368 0.01263 -3.46 0.00055 ***
## UniqueAuthors2 0.11099 0.02202 5.04 4.8e-07 ***
## UniqueAuthors3 0.08743 0.02162 4.04 5.3e-05 ***
## UniqueAuthors4 0.11774 0.02189 5.38 7.7e-08 ***
## UniqueAuthors5 0.13748 0.02137 6.43 1.3e-10 ***
## Year1997 -0.01101 0.02567 -0.43 0.66807
## Year1998 0.00496 0.02493 0.20 0.84240
## Year1999 -0.00982 0.02490 -0.39 0.69328
```

```

## Year2000          0.00861    0.02489    0.35  0.72936
## Year2001          0.02892    0.02536    1.14  0.25415
## Year2002         -0.02286    0.02357   -0.97  0.33204
## Year2003         -0.02524    0.02274   -1.11  0.26706
## Year2004         -0.03403    0.02262   -1.50  0.13254
## Year2005         -0.05896    0.02375   -2.48  0.01305 *
## Year2006         -0.05797    0.02352   -2.46  0.01374 *
## Year2007         -0.00336    0.02370   -0.14  0.88720
## Year2008          0.01583    0.02365    0.67  0.50345
## Year2009         -0.00655    0.02432   -0.27  0.78774
## Year2010          0.02309    0.02400    0.96  0.33605
## Year2011         -0.01282    0.02377   -0.54  0.58967
## Year2012         -0.03228    0.02482   -1.30  0.19344
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.36
## Multiple R-squared:  0.0143, Adjusted R-squared:  0.012
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 3 observations c(715,907,1048) are outliers with |weight| = 0 ( < 1.1e-
05);
## 814 weights are ~= 1. The remaining 8671 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0041 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.05e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1 1.011
## LastAuthorFemale 1.012 1 1.006
## Year 1.030 16 1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19408 -0.24052  0.00238  0.24299  2.04021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.145843   0.017898   64.02  < 2e-16 ***
## FirstAuthorFemale1  0.018510   0.009550    1.94  0.05262 .
## LastAuthorFemale1 -0.041559   0.012614   -3.29  0.00099 ***
## Year1997        -0.018050   0.025731   -0.70  0.48300
## Year1998         0.005187   0.024959    0.21  0.83536
## Year1999        -0.013906   0.024946   -0.56  0.57724
## Year2000         0.005411   0.024980    0.22  0.82852
## Year2001         0.028888   0.025442    1.14  0.25621
## Year2002        -0.021058   0.023611   -0.89  0.37250
## Year2003        -0.019380   0.022712   -0.85  0.39354
## Year2004        -0.030318   0.022646   -1.34  0.18066
## Year2005        -0.055083   0.023919   -2.30  0.02131 *
```

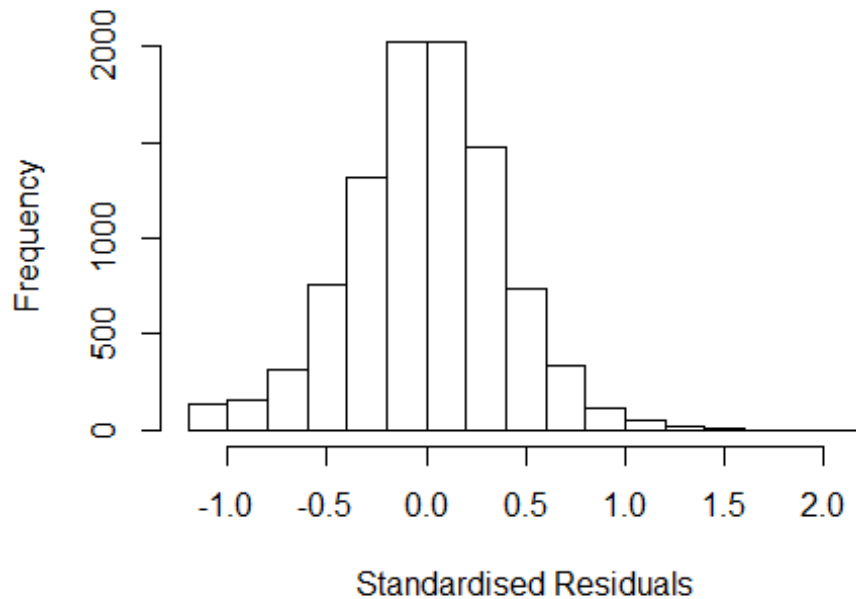
```

## Year2006          -0.053055    0.023623    -2.25    0.02473 *
## Year2007           0.005838    0.023674     0.25    0.80523
## Year2008           0.018776    0.023673     0.79    0.42773
## Year2009          -0.000944    0.024409    -0.04    0.96915
## Year2010           0.029725    0.023984     1.24    0.21523
## Year2011          -0.004629    0.023712    -0.20    0.84524
## Year2012          -0.025907    0.024800    -1.04    0.29622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.00549,    Adjusted R-squared:  0.0036
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(715,907,1048) are outliers with |weight| = 0 ( < 1.1e-
05);
## 801 weights are ~= 1. The remaining 8684 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0028 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.05e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.02 1          1.010
## Year              1.02 16          1.001

```



## Residuals from first author



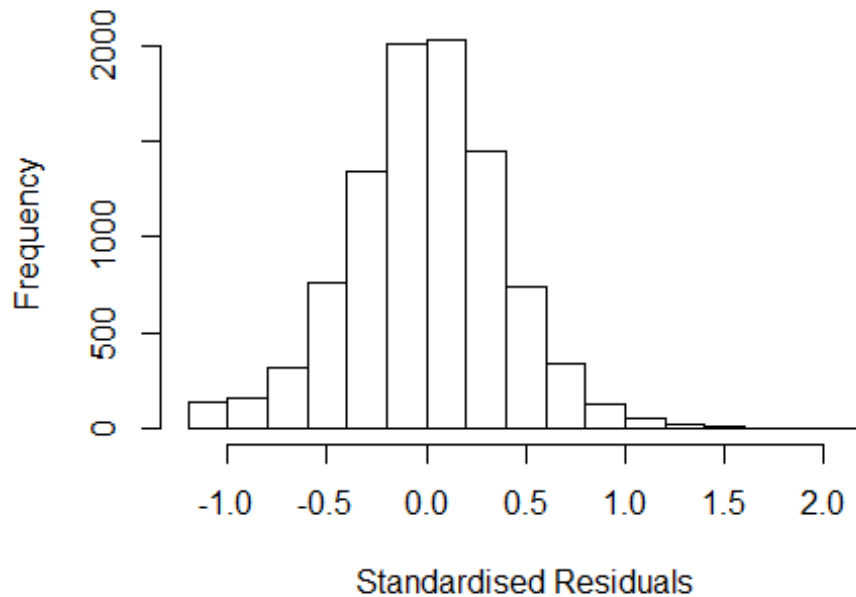
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18774 -0.23964  0.00208  0.24394  2.04391
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.14279    0.01786   63.98  <2e-16 ***
## FirstAuthorFemale1 0.01644    0.00956    1.72   0.085 .
## Year1997        -0.01870    0.02576   -0.73   0.468
## Year1998         0.00538    0.02496    0.22   0.829
## Year1999        -0.01523    0.02494   -0.61   0.542
## Year2000         0.00451    0.02499    0.18   0.857
## Year2001         0.02852    0.02542    1.12   0.262
## Year2002        -0.02199    0.02359   -0.93   0.351
## Year2003        -0.02065    0.02268   -0.91   0.363
## Year2004        -0.03220    0.02262   -1.42   0.155
## Year2005        -0.05587    0.02390   -2.34   0.019 *
## Year2006        -0.05364    0.02359   -2.27   0.023 *
```

```

## Year2007          0.00496    0.02363    0.21    0.834
## Year2008          0.01747    0.02370    0.74    0.461
## Year2009         -0.00151    0.02440   -0.06    0.951
## Year2010          0.02756    0.02393    1.15    0.249
## Year2011         -0.00540    0.02373   -0.23    0.820
## Year2012         -0.02727    0.02479   -1.10    0.271
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.00435,    Adjusted R-squared:  0.00256
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(715,907,1048) are outliers with |weight| = 0 ( < 1.1e-
05);
## 804 weights are ~= 1. The remaining 8681 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0025 0.8640 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.05e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.01 1      1.005
## Year      1.01 16      1.000

```

## Residuals from last author



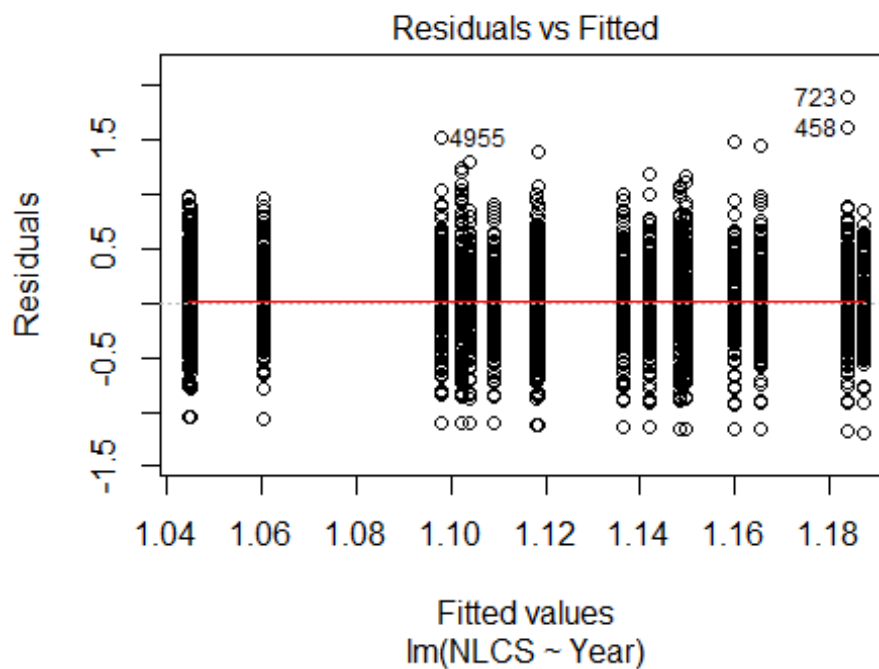
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17980 -0.23921  0.00226  0.24295  2.03759
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.148415   0.017843   64.36  <2e-16 ***
## LastAuthorFemale1 -0.039974   0.012593   -3.17   0.0015 **
## Year1997        -0.018008   0.025739   -0.70   0.4842
## Year1998         0.005274   0.024977    0.21   0.8328
## Year1999        -0.013271   0.024973   -0.53   0.5951
## Year2000         0.005368   0.025013    0.21   0.8301
## Year2001         0.029399   0.025459    1.15   0.2482
## Year2002        -0.020534   0.023606   -0.87   0.3844
## Year2003        -0.018279   0.022719   -0.80   0.4211
## Year2004        -0.029317   0.022642   -1.29   0.1954
## Year2005        -0.053802   0.023897   -2.25   0.0244 *
## Year2006        -0.051892   0.023622   -2.20   0.0281 *
```

```

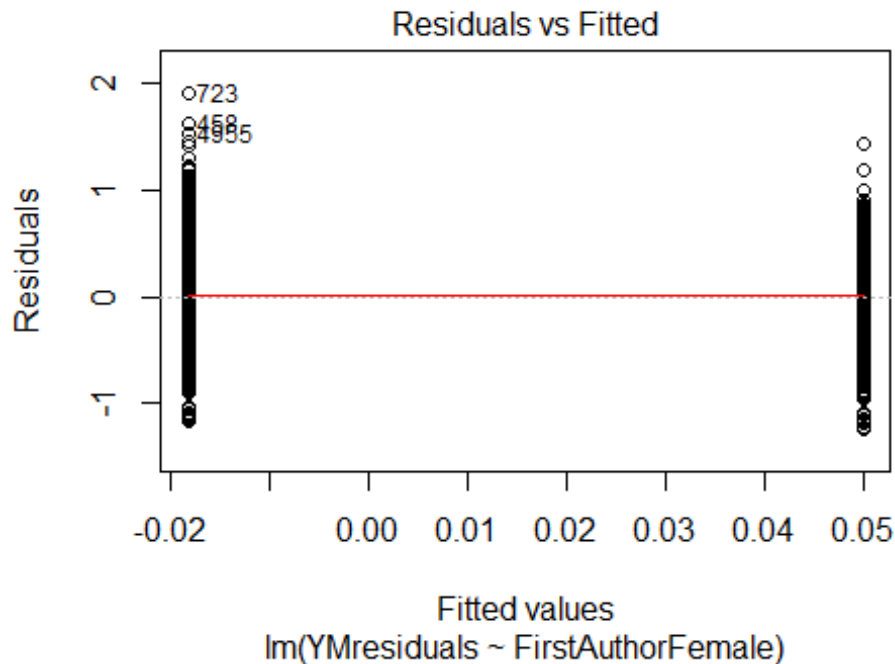
## Year2007      0.007442  0.023660  0.31  0.7531
## Year2008      0.020642  0.023658  0.87  0.3830
## Year2009      0.000671  0.024396  0.03  0.9781
## Year2010      0.031384  0.023958  1.31  0.1902
## Year2011     -0.002481  0.023671 -0.10  0.9165
## Year2012     -0.024188  0.024764 -0.98  0.3287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.00511,    Adjusted R-squared:  0.00333
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(715,907,1048) are outliers with |weight| = 0 ( < 1.1e-
05);
## 824 weights are ~ = 1. The remaining 8661 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0031 0.8650 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.05e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 9488"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1607"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 305 373 296 312 322 265 321 252 290 231 271 241 257 319 283
## 2011 2012
## 292 252
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 162 183 143 167 170 115 193 162 195 147 163 161 167 223 195
## 2011 2012
## 212 184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 149 165 137 148 144 103 162 146 167 126 149 147 148 194 171
## 2011 2012
## 187 164
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.04
```

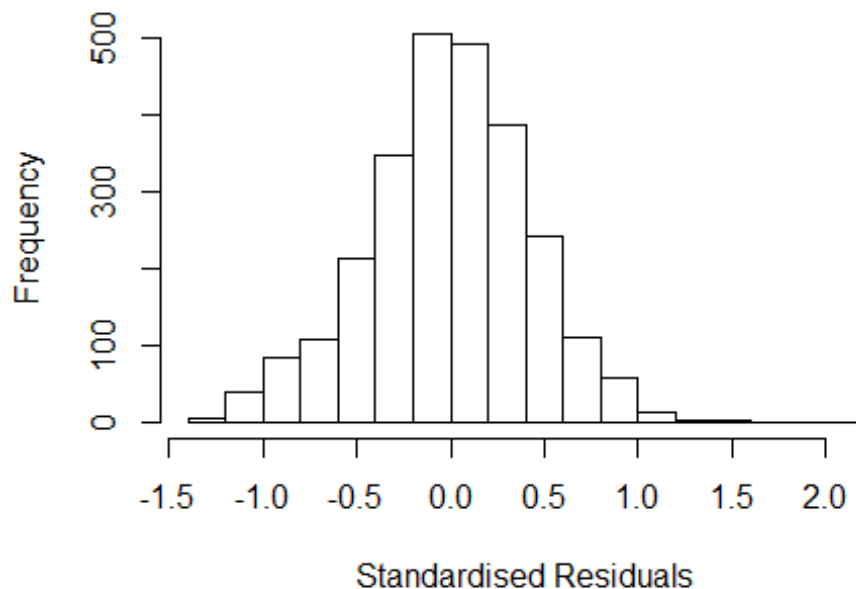


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 27, df = 1, p-value = 2e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.26413647726631"
## [1] "Male first author team size 2018 geometric mean: 3.64736800445758"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.79798135688208"
## [1] "Male last author team size 2018 geometric mean: 3.89321359143814"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.057 1 1.028
## LastAuthorFemale 1.048 1 1.024
## UniqueAuthors 1.264 4 1.030
## Year 1.324 16 1.009
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24563 -0.26617  0.00217  0.27601  2.08567
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94515    0.05588   16.91 < 2e-16 ***
## FirstAuthorFemale1 0.04538    0.01795    2.53  0.0115 *
## LastAuthorFemale1 -0.00311    0.02465   -0.13  0.8996
## UniqueAuthors2    0.19905    0.04488    4.43 9.6e-06 ***
## UniqueAuthors3    0.23686    0.04343    5.45 5.4e-08 ***
## UniqueAuthors4    0.27421    0.04488    6.11 1.1e-09 ***
## UniqueAuthors5    0.27141    0.04331    6.27 4.3e-10 ***
## Year1997          0.04318    0.05444    0.79  0.4277
## Year1998          0.01644    0.05735    0.29  0.7743
## Year1999         -0.04907    0.05296   -0.93  0.3543
```

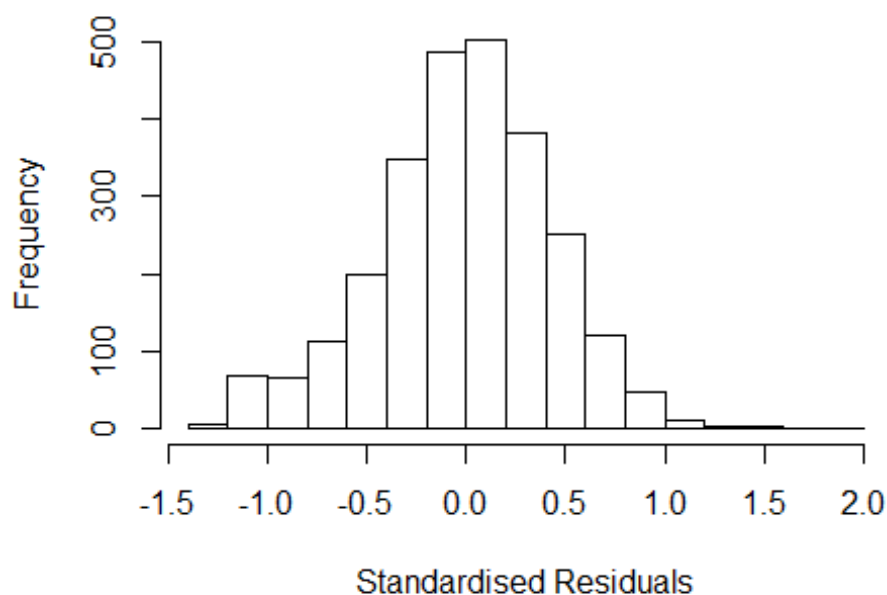
```

## Year2000      0.02627    0.06108    0.43    0.6671
## Year2001     -0.00840    0.05548   -0.15    0.8796
## Year2002     -0.07073    0.05371   -1.32    0.1880
## Year2003     -0.08240    0.05195   -1.59    0.1128
## Year2004     -0.02172    0.05417   -0.40    0.6885
## Year2005     -0.04797    0.05588   -0.86    0.3908
## Year2006     -0.02241    0.05684   -0.39    0.6934
## Year2007     -0.04842    0.05446   -0.89    0.3740
## Year2008      0.01004    0.05086    0.20    0.8435
## Year2009     -0.09557    0.05091   -1.88    0.0606 .
## Year2010     -0.10556    0.05140   -2.05    0.0401 *
## Year2011     -0.14272    0.05290   -2.70    0.0070 **
## Year2012     -0.15116    0.05338   -2.83    0.0047 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.0473, Adjusted R-squared:  0.0392
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 277 is an outlier with |weight| = 0 ( < 3.8e-05);
## 223 weights are ~= 1. The remaining 2383 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0872 0.8670 0.9520 0.8990 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1 1.023
## LastAuthorFemale 1.036 1 1.018
## Year 1.081 16 1.002

```



## Residuals from first and last author



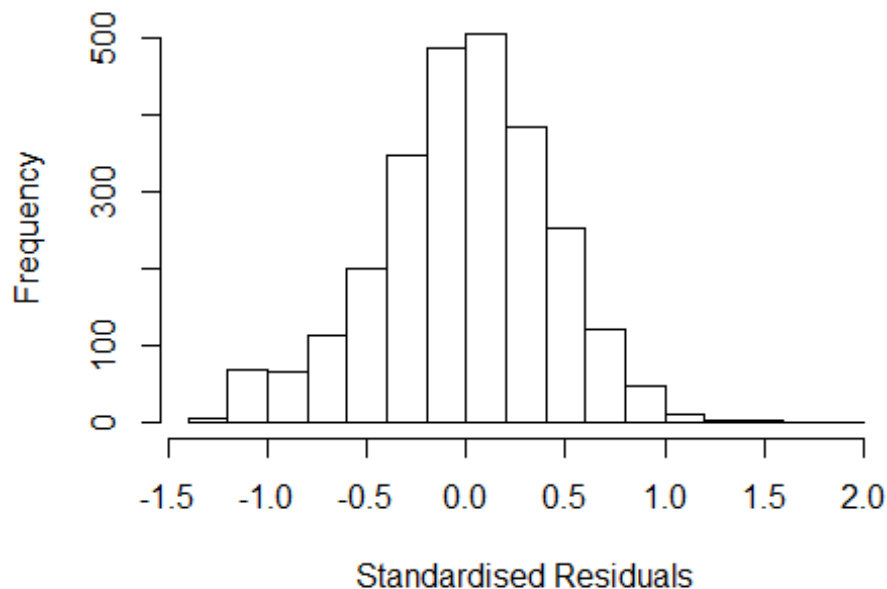
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24433 -0.27029 0.00771 0.28176 1.89596
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14486 0.04111 27.85 < 2e-16 ***
## FirstAuthorFemale1 0.06628 0.01804 3.67 0.00024 ***
## LastAuthorFemale1 -0.00135 0.02464 -0.05 0.95647
## Year1997 0.03318 0.05480 0.61 0.54494
## Year1998 0.02415 0.05769 0.42 0.67550
## Year1999 -0.03695 0.05332 -0.69 0.48842
## Year2000 0.03145 0.06161 0.51 0.60978
## Year2001 0.00592 0.05639 0.10 0.91641
## Year2002 -0.04924 0.05456 -0.90 0.36692
## Year2003 -0.05325 0.05227 -1.02 0.30850
## Year2004 0.00543 0.05372 0.10 0.91953
## Year2005 -0.04175 0.05555 -0.75 0.45230
```

```

## Year2006          0.00774    0.05673    0.14  0.89154
## Year2007          -0.02482    0.05417   -0.46  0.64686
## Year2008           0.03658    0.05046    0.72  0.46861
## Year2009          -0.06390    0.05080   -1.26  0.20854
## Year2010          -0.07624    0.05125   -1.49  0.13699
## Year2011          -0.10724    0.05274   -2.03  0.04211 *
## Year2012          -0.11225    0.05298   -2.12  0.03421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.00904
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 201 weights are ~= 1. The remaining 2406 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0032 0.8680 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1      1.023
## Year              1.046 16      1.001

```

## Residuals from first author



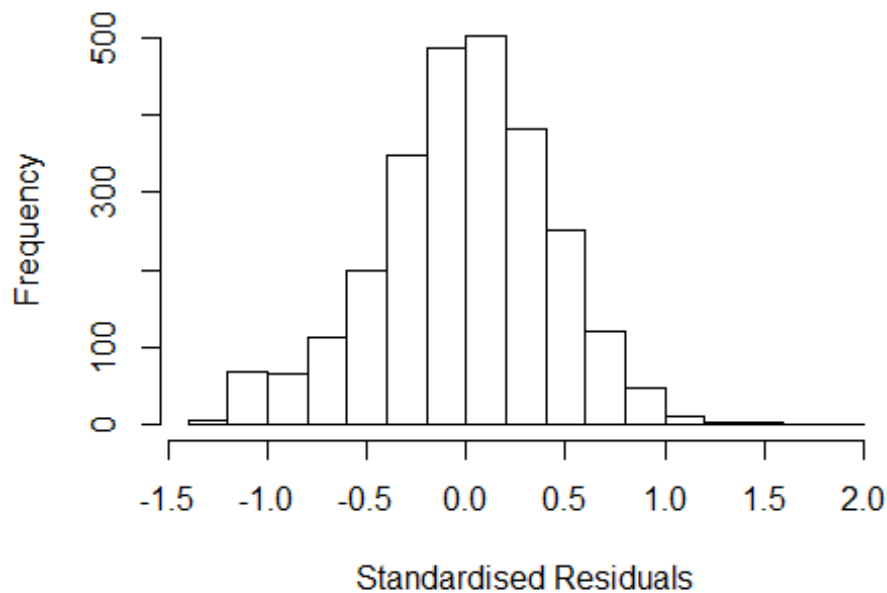
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24419 -0.27019  0.00781  0.28190  1.89601
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14480    0.04109   27.86 < 2e-16 ***
## FirstAuthorFemale1 0.06621    0.01804    3.67 0.00025 ***
## Year1997         0.03318    0.05482    0.61 0.54499
## Year1998         0.02403    0.05762    0.42 0.67669
## Year1999        -0.03702    0.05330   -0.69 0.48736
## Year2000         0.03138    0.06158    0.51 0.61035
## Year2001         0.00595    0.05639    0.11 0.91594
## Year2002        -0.04934    0.05456   -0.90 0.36592
## Year2003        -0.05337    0.05229   -1.02 0.30753
## Year2004         0.00538    0.05371    0.10 0.92017
## Year2005        -0.04184    0.05549   -0.75 0.45085
## Year2006         0.00764    0.05667    0.13 0.89277
```

```

## Year2007          -0.02498    0.05409   -0.46  0.64419
## Year2008           0.03641    0.05031    0.72  0.46940
## Year2009          -0.06400    0.05077   -1.26  0.20761
## Year2010          -0.07636    0.05124   -1.49  0.13630
## Year2011          -0.10731    0.05271   -2.04  0.04186 *
## Year2012          -0.11230    0.05300   -2.12  0.03421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.00942
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2407 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0031 0.8680 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1          1.017
## Year            1.035 16          1.001

```

## Residuals from last author



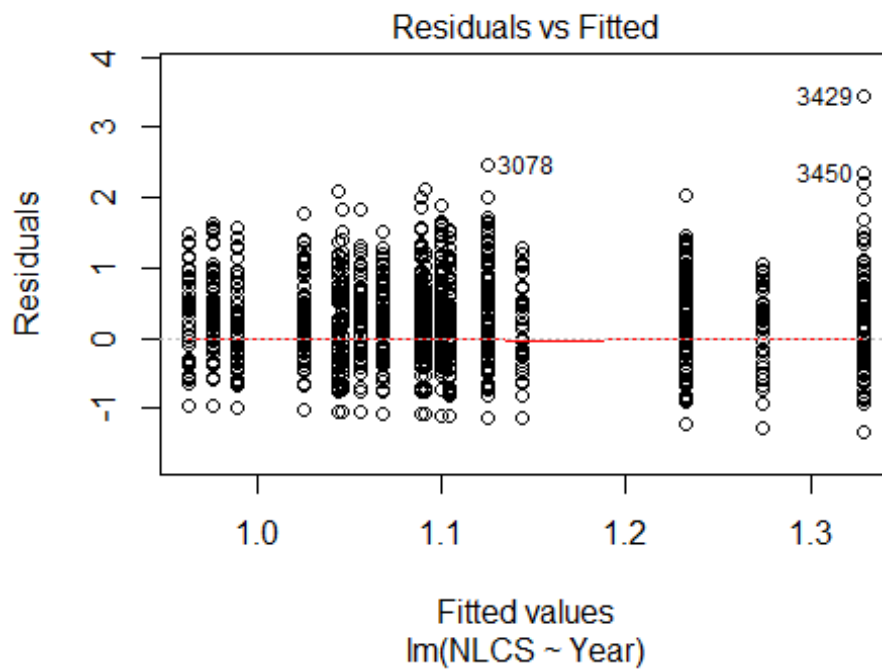
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20554 -0.27009  0.00524  0.28372  1.88504
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.15868    0.04111   28.18  <2e-16 ***
## LastAuthorFemale1 0.00484    0.02468    0.20   0.844
## Year1997        0.03028    0.05514    0.55   0.583
## Year1998        0.01957    0.05795    0.34   0.736
## Year1999       -0.03609    0.05348   -0.67   0.500
## Year2000        0.03244    0.06197    0.52   0.601
## Year2001        0.00641    0.05648    0.11   0.910
## Year2002       -0.04498    0.05450   -0.83   0.409
## Year2003       -0.04653    0.05225   -0.89   0.373
## Year2004        0.01096    0.05398    0.20   0.839
## Year2005       -0.03161    0.05556   -0.57   0.569
## Year2006        0.01498    0.05683    0.26   0.792
```

```

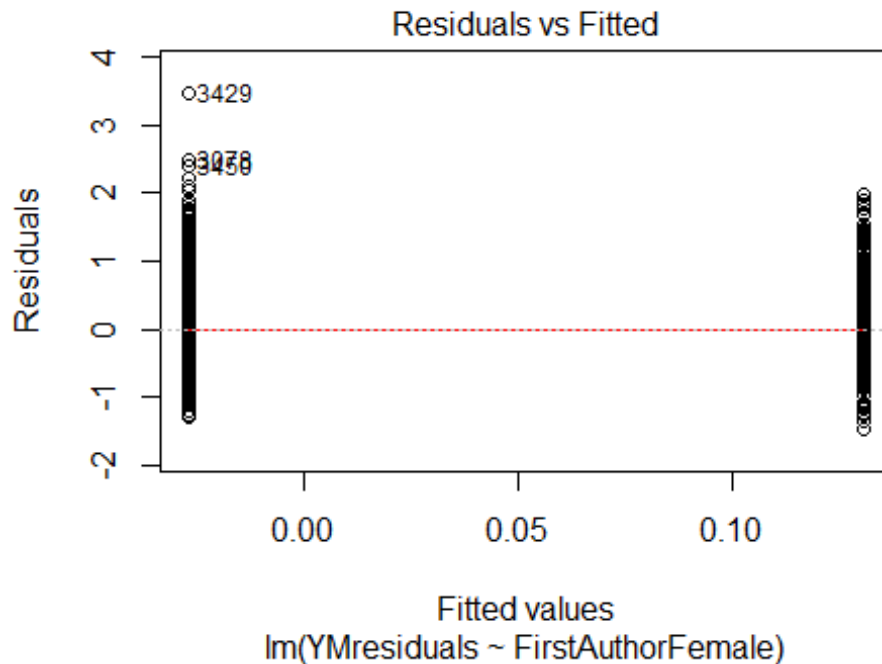
## Year2007          -0.02092      0.05450   -0.38      0.701
## Year2008           0.04686      0.05075     0.92      0.356
## Year2009          -0.05700      0.05098    -1.12      0.264
## Year2010          -0.06836      0.05155    -1.33      0.185
## Year2011          -0.09872      0.05291    -1.87      0.062 .
## Year2012          -0.10497      0.05322    -1.97      0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.0111, Adjusted R-squared:  0.00464
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2422 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0048 0.8710 0.9530 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2607"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1700"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 185 110 157 148 157 129 135 81 76 128 167 205 235 262 251
## 2011 2012
## 269 203
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 72 82 83 101 74 84 48 43 99 108 137 168 181 166
## 2011 2012

```

```
## 177 145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 91 70 73 81 94 72 71 43 40 91 99 118 145 158 151
## 2011 2012
## 160 121
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.004
```



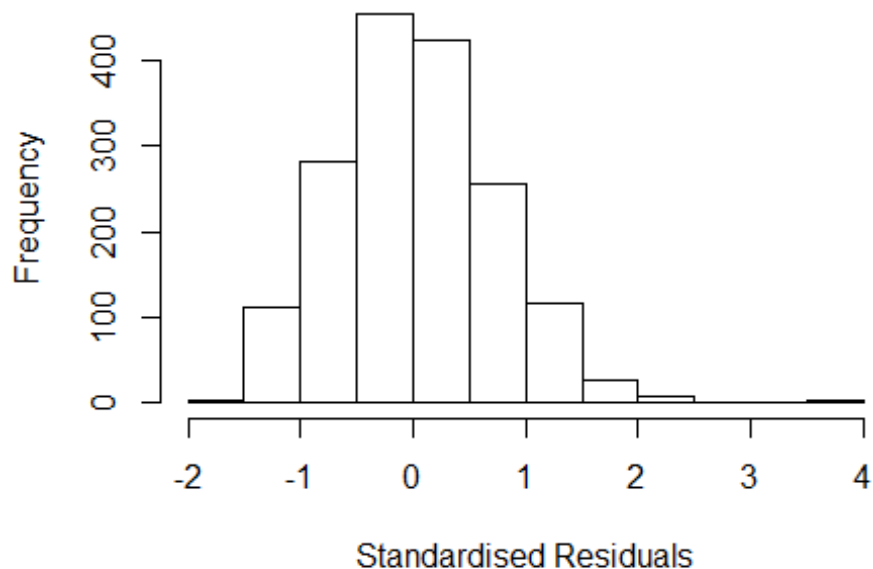
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.21, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 2.30631543327954"
## [1] "Male first author team size 2018 geometric mean: 2.2882210987398"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.2808060129244"
## [1] "Male last author team size 2018 geometric mean: 2.29330266566903"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.227 1          1.108
## LastAuthorFemale  1.204 1          1.097
## UniqueAuthors    1.129 4          1.015
## Year              1.175 16         1.005
```



## Residuals from first and last author and team size



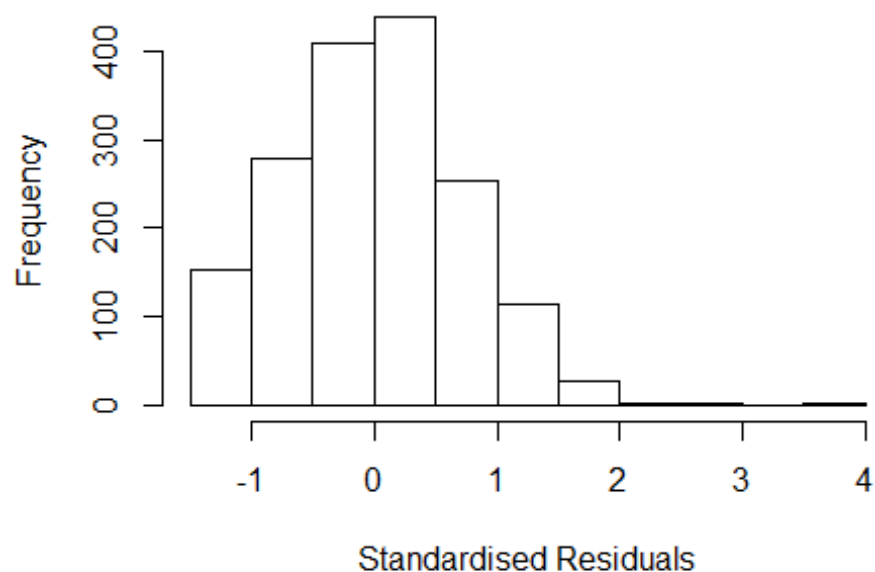
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3429 84864614709 4.765 2012    1700      2    3.507
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5230 -0.4737 -0.0104  0.4801  3.5065
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.80451    0.07969   10.10 < 2e-16 ***
## FirstAuthorFemale1 0.08549    0.05113    1.67  0.095 .
## LastAuthorFemale1 0.04630    0.05196    0.89  0.373
## UniqueAuthors2    0.23063    0.04164    5.54 3.5e-08 ***
## UniqueAuthors3    0.25377    0.04994    5.08 4.2e-07 ***
## UniqueAuthors4    0.43324    0.08629    5.02 5.7e-07 ***
## UniqueAuthors5    0.52448    0.07273    7.21 8.4e-13 ***
## Year1997          0.00930    0.12816    0.07  0.942
## Year1998          0.03217    0.12216    0.26  0.792
## Year1999          0.12338    0.10396    1.19  0.236
```

```

## Year2000      0.06851      0.10246      0.67      0.504
## Year2001      0.02080      0.11349      0.18      0.855
## Year2002     -0.01190      0.11569     -0.10      0.918
## Year2003      0.34296      0.13830      2.48      0.013 *
## Year2004      0.11910      0.13195      0.90      0.367
## Year2005      0.09042      0.10463      0.86      0.388
## Year2006      0.04722      0.09714      0.49      0.627
## Year2007      0.00577      0.09722      0.06      0.953
## Year2008      0.07914      0.09244      0.86      0.392
## Year2009      0.06656      0.09244      0.72      0.472
## Year2010      0.19396      0.09956      1.95      0.052 .
## Year2011      0.09005      0.10403      0.87      0.387
## Year2012      0.20020      0.10503      1.91      0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.0645, Adjusted R-squared:  0.052
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1582 is an outlier with |weight| = 0 ( < 6e-05);
## 149 weights are ~= 1. The remaining 1528 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.207  0.867  0.951  0.913  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.96e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.193 1           1.092
## LastAuthorFemale  1.175 1           1.084
## Year              1.076 16           1.002

```

## Residuals from first and last author



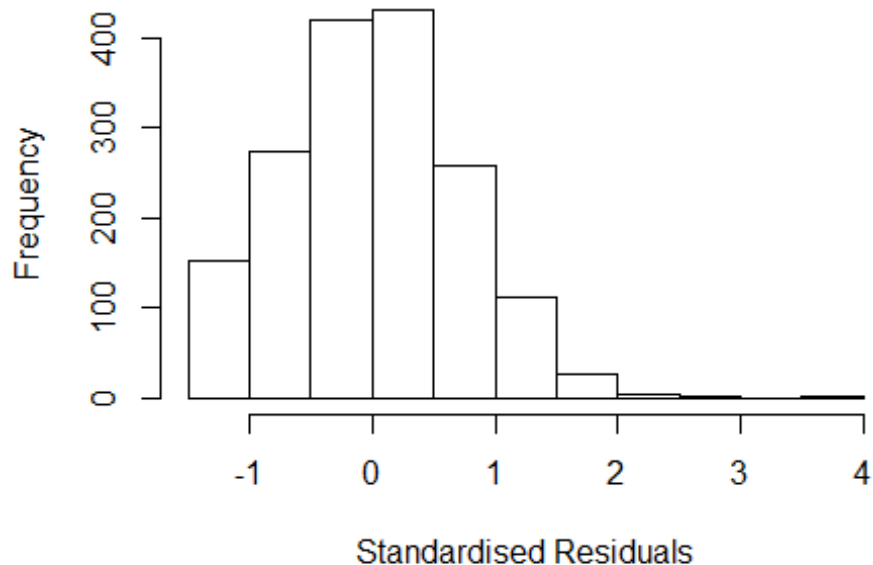
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3078 79960204163 3.584 2011      1700      2      2.525
## 3429 84864614709 4.765 2012      1700      2      3.582
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.37115 -0.52534 -0.00212  0.47307  3.58217
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9367    0.0798   11.74  <2e-16 ***
## FirstAuthorFemale1 0.1217    0.0518    2.35  0.0189 *
## LastAuthorFemale1 0.0667    0.0540    1.23  0.2172
## Year1997       -0.0181    0.1268   -0.14  0.8863
## Year1998        0.0157    0.1258    0.12  0.9006
## Year1999        0.1268    0.1087    1.17  0.2436
## Year2000        0.0865    0.1055    0.82  0.4125
## Year2001        0.0365    0.1138    0.32  0.7483
## Year2002        0.0457    0.1159    0.39  0.6933
## Year2003        0.3588    0.1384    2.59  0.0096 **
## Year2004        0.1717    0.1386    1.24  0.2156
```

```

## Year2005          0.1276      0.1087      1.17      0.2404
## Year2006          0.0626      0.1036      0.60      0.5459
## Year2007          0.0243      0.1016      0.24      0.8114
## Year2008          0.1080      0.0953      1.13      0.2573
## Year2009          0.1084      0.0955      1.14      0.2563
## Year2010          0.2167      0.1043      2.08      0.0378 *
## Year2011          0.1218      0.1076      1.13      0.2576
## Year2012          0.2462      0.1098      2.24      0.0251 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.737
## Multiple R-squared:  0.0199, Adjusted R-squared:  0.00925
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1582 is an outlier with |weight| = 0 ( < 6e-05);
## 144 weights are ~= 1. The remaining 1533 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.216  0.860  0.949   0.915  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.045 1      1.022
## Year              1.045 16      1.001

```

## Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3078 79960204163 3.584 2011      1700      2      2.525
## 3429 84864614709 4.765 2012      1700      2      3.582
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.33708 -0.51681 -0.00351  0.47790  3.57489
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9401     0.0803   11.71  <2e-16 ***
## FirstAuthorFemale1 0.1470     0.0482    3.05  0.0024 **
## Year1997        -0.0143     0.1270   -0.11  0.9101
## Year1998         0.0225     0.1255    0.18  0.8575
## Year1999         0.1292     0.1089    1.19  0.2359
## Year2000         0.0880     0.1058    0.83  0.4054
## Year2001         0.0370     0.1140    0.32  0.7458
## Year2002         0.0484     0.1158    0.42  0.6758
## Year2003         0.3579     0.1395    2.57  0.0104 *
## Year2004         0.1718     0.1400    1.23  0.2201
## Year2005         0.1314     0.1086    1.21  0.2266
```

```

## Year2006          0.0647      0.1040      0.62      0.5336
## Year2007          0.0316      0.1018      0.31      0.7563
## Year2008          0.1104      0.0959      1.15      0.2497
## Year2009          0.1115      0.0958      1.16      0.2445
## Year2010          0.2187      0.1045      2.09      0.0366 *
## Year2011          0.1227      0.1079      1.14      0.2555
## Year2012          0.2500      0.1099      2.27      0.0231 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.738
## Multiple R-squared:  0.0189, Adjusted R-squared:  0.00887
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1582 is an outlier with |weight| = 0 ( < 6e-05);
## 139 weights are ~= 1. The remaining 1538 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.219  0.862  0.951   0.916   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.03 1          1.015
## Year          1.03 16          1.001

## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3078 79960204163 3.584 2011      1700      2      2.525
## 3429 84864614709 4.765 2012      1700      2      3.582
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"

```

```

## Residuals:
##      Min      1Q    Median      3Q      Max
## -1.31e+00 -5.14e-01  5.96e-05  4.70e-01  3.57e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9480    0.0797   11.90  <2e-16 ***
## LastAuthorFemale1 0.1157    0.0502    2.31  0.0212 *
## Year1997       -0.0228    0.1269   -0.18  0.8571
## Year1998        0.0305    0.1261    0.24  0.8091
## Year1999        0.1268    0.1086    1.17  0.2430
## Year2000        0.0906    0.1059    0.86  0.3920
## Year2001        0.0439    0.1148    0.38  0.7022
## Year2002        0.0384    0.1164    0.33  0.7418
## Year2003        0.3612    0.1371    2.63  0.0085 **
## Year2004        0.1811    0.1368    1.32  0.1858
## Year2005        0.1241    0.1090    1.14  0.2554
## Year2006        0.0629    0.1038    0.61  0.5443
## Year2007        0.0205    0.1015    0.20  0.8400
## Year2008        0.1026    0.0953    1.08  0.2819
## Year2009        0.1129    0.0956    1.18  0.2376
## Year2010        0.2206    0.1045    2.11  0.0349 *
## Year2011        0.1293    0.1076    1.20  0.2299
## Year2012        0.2478    0.1094    2.27  0.0236 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.738
## Multiple R-squared:  0.0165, Adjusted R-squared:  0.00645
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1582 is an outlier with |weight| = 0 ( < 6e-05);
## 146 weights are ~ = 1. The remaining 1531 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.225  0.862  0.950  0.915  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.96e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats

```

```

##                                "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1678"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1701"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1     6     2    15     7     3     1     6    22    44    52    25    29    33    40
## 2012
##    33
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1     4     2    10     5     2     1     4    13    28    34    18    22    23    28
## 2012
##    21
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1     4     2    10     5     2     1     4    12    25    29    17    20    19    27
## 2012
##    17
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.79931528956993"
## [1] "Male first author team size 2018 geometric mean: 1.78685365452093"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 88, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.64754897244207"
## [1] "Male last author team size 2018 geometric mean: 1.83544597703591"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

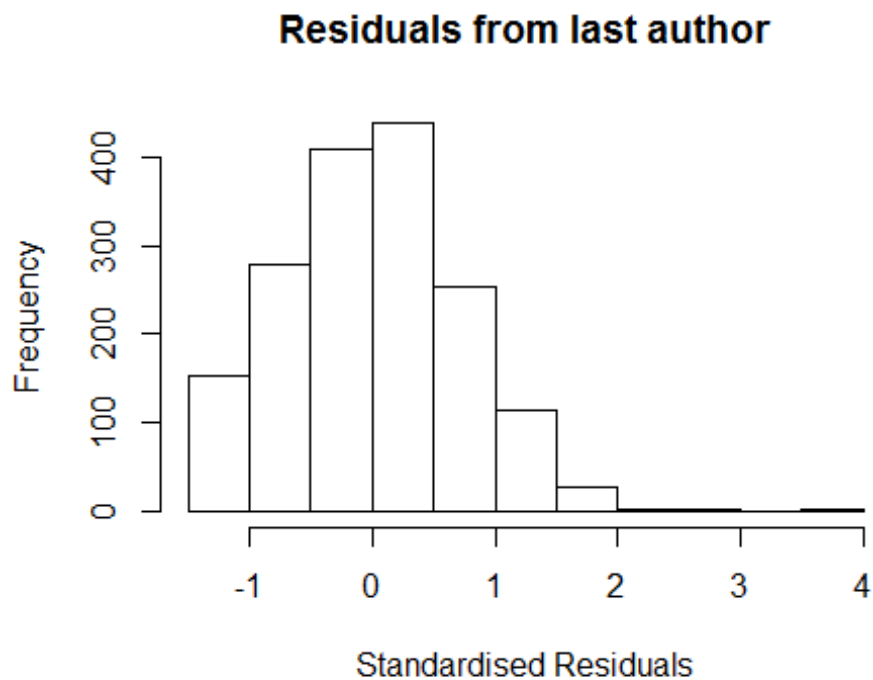
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 54, p-value = 0.6

```



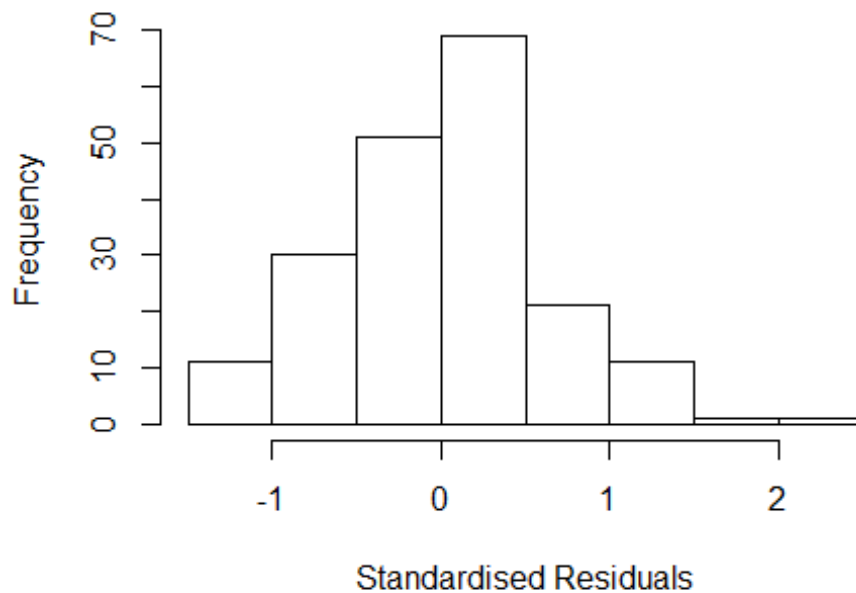
```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```



##		GVIF	Df	$GVIF^{(1/(2*Df))}$
##	FirstAuthorFemale	2.724e+12	1	1.650e+06
##	LastAuthorFemale	2.456e+00	1	1.567e+00
##	UniqueAuthors	6.952e+12	4	4.030e+01
##	Year	3.546e+13	15	2.829e+00

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3878 -0.3548 0.0383 0.3600 2.0964
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5960 0.0000 Inf < 2e-16 ***
## FirstAuthorFemale1 -0.1697 0.1520 -1.12 0.26583
## LastAuthorFemale1 0.3273 0.1596 2.05 0.04183 *
## UniqueAuthors2 0.3013 0.1344 2.24 0.02626 *
## UniqueAuthors3 0.2103 0.1258 1.67 0.09623 .
## UniqueAuthors4 0.1292 0.2069 0.62 0.53313
## UniqueAuthors5 -0.1764 0.1719 -1.03 0.30644
## Year1998 -0.2879 0.1614 -1.78 0.07630 .
## Year1999 0.6659 0.1128 5.91 1.8e-08 ***
## Year2000 0.2674 0.1719 1.55 0.12181
```

```

## Year2001          0.2002      0.2840      0.71  0.48172
## Year2002          0.1539      0.2840      0.54  0.58869
## Year2003          0.9515      0.2053      4.63  7.1e-06 ***
## Year2004          0.0483      0.4630      0.10  0.91699
## Year2005         -0.0601      0.1961     -0.31  0.75961
## Year2006          0.4905      0.1606      3.05  0.00261 **
## Year2007          0.1605      0.1280      1.25  0.21149
## Year2008          0.2995      0.1634      1.83  0.06853 .
## Year2009          0.4047      0.1637      2.47  0.01436 *
## Year2010          0.4446      0.1318      3.37  0.00092 ***
## Year2011          0.3142      0.1572      2.00  0.04720 *
## Year2012          0.2207      0.1780      1.24  0.21655
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.644
## Multiple R-squared:  0.151, Adjusted R-squared:  0.0482
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 177 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.268  0.878  0.968  0.914  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 5.346e+12  1      2.312e+06
## LastAuthorFemale  2.264e+00  1      1.505e+00
## Year              8.194e+12 15      2.694e+00
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##

```

```

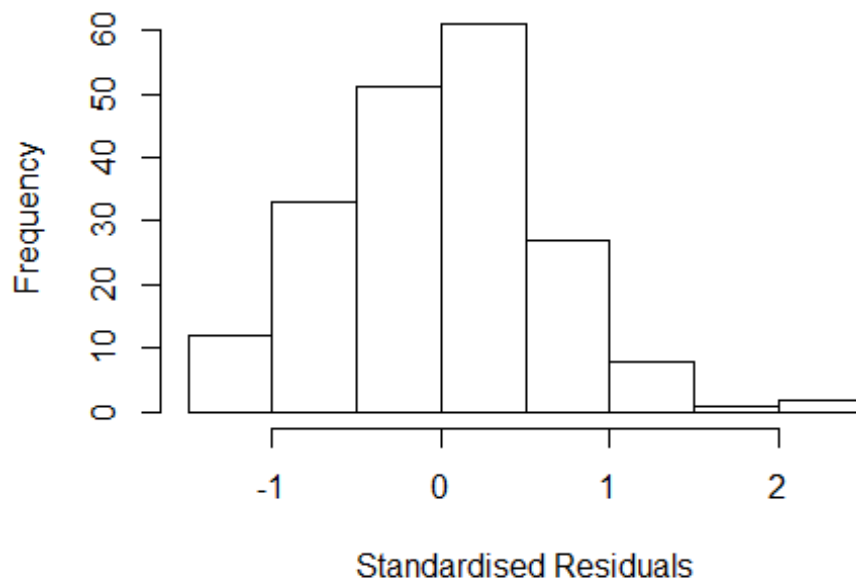
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24595 -0.45867  0.00598  0.40560  2.13722
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    5.96e-01   2.92e-08  2.04e+07 < 2e-16 ***
## FirstAuthorFemale1 -1.90e-01   1.71e-01 -1.11e+00  0.26885
## LastAuthorFemale1  3.48e-01   1.73e-01  2.01e+00  0.04619 *
## Year1998        -2.88e-01   1.61e-01 -1.79e+00  0.07578 .
## Year1999         8.16e-01   1.66e-02  4.91e+01 < 2e-16 ***
## Year2000         3.01e-01   1.62e-01  1.86e+00  0.06520 .
## Year2001         2.73e-01   2.85e-01  9.60e-01  0.33955
## Year2002         3.04e-01   4.14e-01  7.40e-01  0.46308
## Year2003         1.27e+00   1.71e-01  7.44e+00  4.2e-12 ***
## Year2004         2.62e-01   3.86e-01  6.80e-01  0.49853
## Year2005         4.65e-02   1.72e-01  2.70e-01  0.78731
## Year2006         6.09e-01   1.43e-01  4.26e+00  3.4e-05 ***
## Year2007         2.69e-01   1.23e-01  2.20e+00  0.02942 *
## Year2008         5.30e-01   1.31e-01  4.05e+00  7.7e-05 ***
## Year2009         5.20e-01   1.68e-01  3.09e+00  0.00230 **
## Year2010         5.23e-01   1.33e-01  3.94e+00  0.00012 ***
## Year2011         4.26e-01   1.65e-01  2.58e+00  0.01083 *
## Year2012         2.80e-01   1.67e-01  1.68e+00  0.09475 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.657
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0285
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 178 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.269  0.885  0.954  0.914  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      5.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd

```

```
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```

### Residuals from first and last author



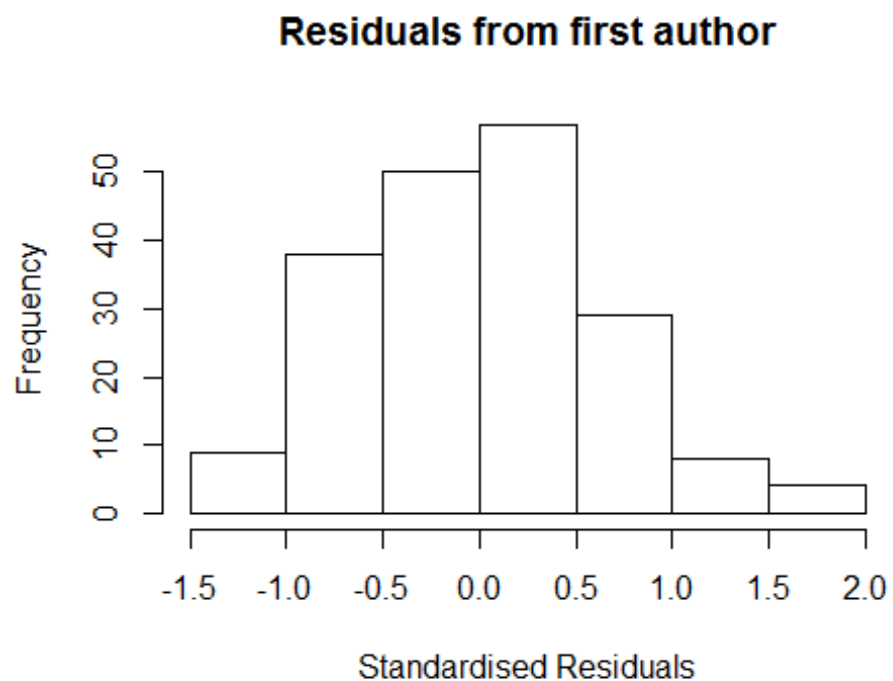
```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.921e+13 1 4.382e+06
## Year 1.921e+13 15 2.772e+00

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2051 -0.4812 0.0129 0.3869 1.9858
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.596000   0.000000    Inf < 2e-16 ***
## FirstAuthorFemale1 0.000994   0.139784    0.01  0.99433
## Year1998       -0.287880   0.160551   -1.79  0.07466 .
## Year1999        0.816500   0.016621   49.13 < 2e-16 ***
## Year2000        0.300119   0.161280    1.86  0.06441 .
## Year2001        0.359821   0.276660    1.30  0.19508
## Year2002        0.304500   0.408987    0.74  0.45754
## Year2003        1.082006   0.139784    7.74  7.2e-13 ***
## Year2004        0.287025   0.345765    0.83  0.40759
## Year2005        0.063548   0.185398    0.34  0.73218
## Year2006        0.609052   0.134807    4.52  1.1e-05 ***
## Year2007        0.340996   0.127208    2.68  0.00804 **
## Year2008        0.581101   0.145390    4.00  9.4e-05 ***
## Year2009        0.528267   0.175112    3.02  0.00293 **
## Year2010        0.555226   0.140291    3.96  0.00011 ***
## Year2011        0.471304   0.155674    3.03  0.00283 **
## Year2012        0.296687   0.162021    1.83  0.06875 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.69
## Multiple R-squared:  0.0827, Adjusted R-squared:  0.000217
## Convergence in 9 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.387  0.892  0.956  0.922  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00        5.00e-01      4.69e+00      1.00e-07
##      rel.tol        solve.tol      eps.outlier      eps.x
##      1.00e-07        1.00e-07      5.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01        5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
```

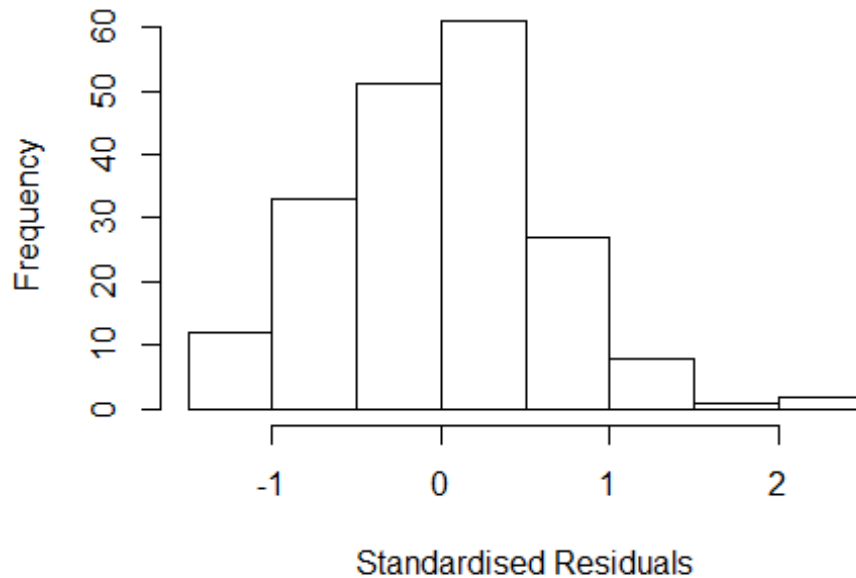
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in cov2cor(v): diag(.) had 0 or NA entries; non-finite result is
## doubtful
```



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale  NaN  1          NaN
## Year              NaN 15          NaN
```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1927 -0.4675 0.0123 0.3541 2.0476
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5960 0.0000 Inf < 2e-16 ***
## LastAuthorFemale1 0.2536 0.1446 1.75 0.08109 .
## Year1998 -0.2879 0.1607 -1.79 0.07490 .
## Year1999 0.8165 0.0166 49.13 < 2e-16 ***
## Year2000 0.3003 0.1615 1.86 0.06457 .
## Year2001 0.2150 0.2857 0.75 0.45257
## Year2002 0.3045 0.4101 0.74 0.45874
## Year2003 1.0830 0.0000 Inf < 2e-16 ***
## Year2004 0.2474 0.4067 0.61 0.54368
## Year2005 0.0364 0.1793 0.20 0.83921
## Year2006 0.5967 0.1415 4.22 3.9e-05 ***
## Year2007 0.2845 0.1238 2.30 0.02274 *
```

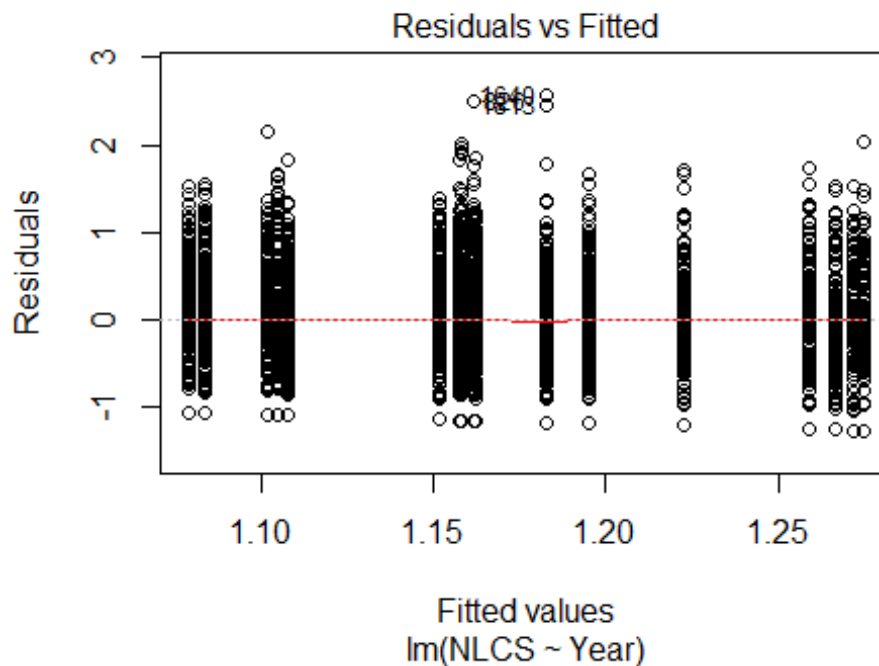


```

## Year2008          0.5291      0.1285      4.12  5.8e-05 ***
## Year2009          0.4918      0.1679      2.93  0.00384 **
## Year2010          0.4934      0.1317      3.75  0.00024 ***
## Year2011          0.4029      0.1618      2.49  0.01370 *
## Year2012          0.2576      0.1627      1.58  0.11518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.682
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0211
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 179 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.348  0.890  0.957  0.920  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 195"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1702"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  239  254  292  293  293  309  224  222  223  251  334  328  328  363  329
## 2011 2012
##  326  344
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  134  144  177  162  161  167  142  157  134  175  222  222  186  253  228
## 2011 2012
##  224  248

```

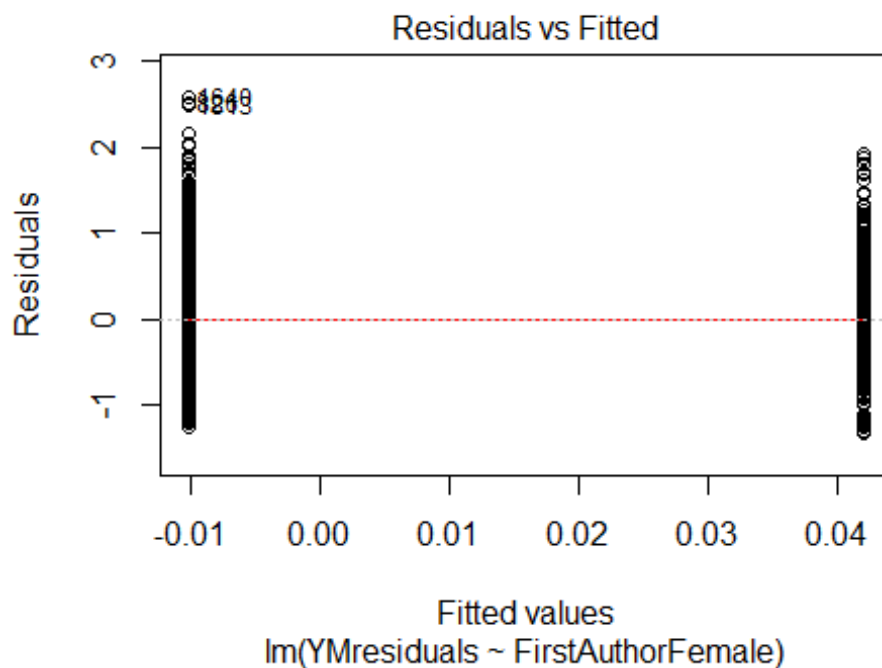
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 122 130 162 146 150 154 127 140 121 150 196 194 158 226 198
## 2011 2012
## 210 218
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.005
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.37, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 2.54331010661263"
## [1] "Male first author team size 2018 geometric mean: 2.0959033315161"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
```

```
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.63214802590498"
## [1] "Male last author team size 2018 geometric mean: 2.11483758744648"

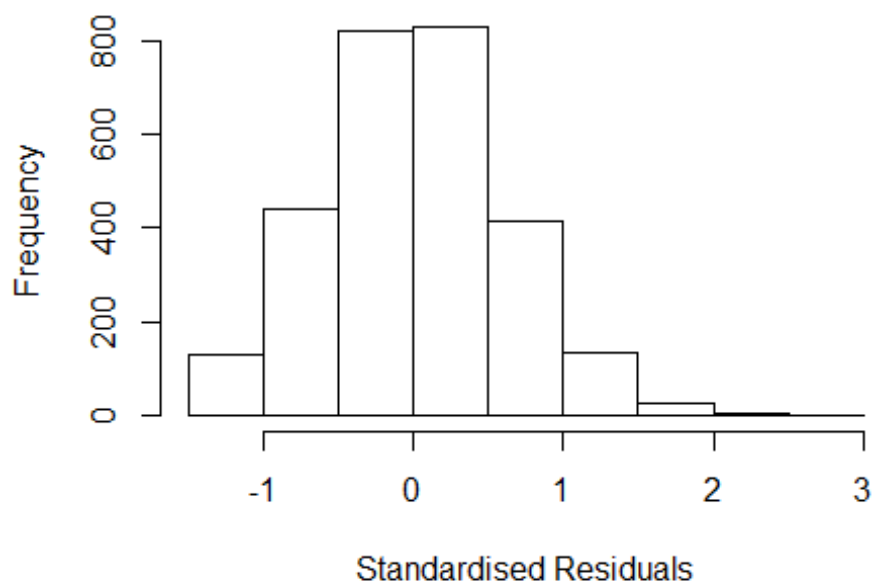
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 160, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.222	1	1.105
LastAuthorFemale	1.213	1	1.101
UniqueAuthors	1.184	4	1.021
Year	1.195	16	1.006

## Residuals from first and last author and team size



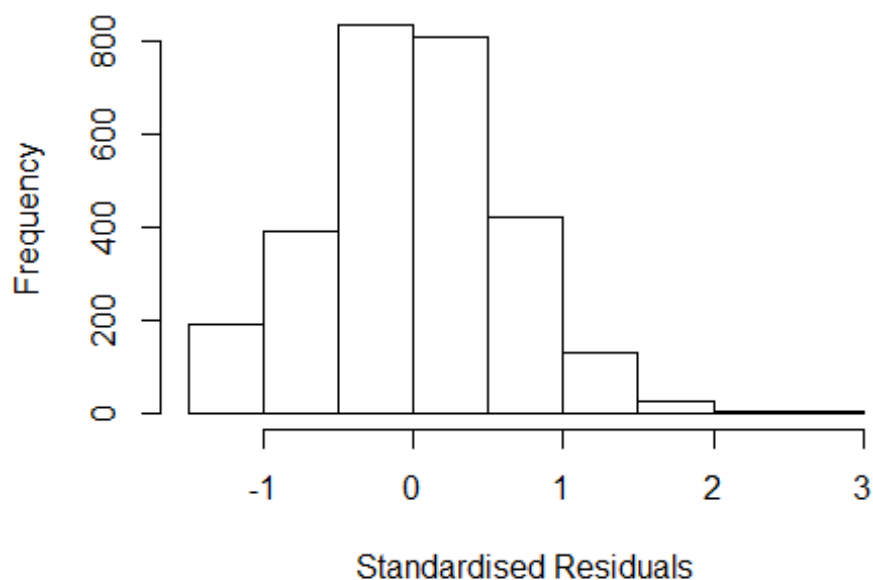
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1640 0001224048 3.749 2001      1702      4      2.759
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45962 -0.40308  0.00325  0.39721  2.75877
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.99480    0.07241   13.74 < 2e-16 ***
## FirstAuthorFemale1 0.01967    0.03241    0.61  0.544
## LastAuthorFemale1 0.08584    0.03449    2.49  0.013 *
## UniqueAuthors2    0.18105    0.02880    6.29 3.8e-10 ***
## UniqueAuthors3    0.19748    0.03382    5.84 5.8e-09 ***
## UniqueAuthors4    0.32960    0.05126    6.43 1.5e-10 ***
## UniqueAuthors5    0.27038    0.05784    4.67 3.1e-06 ***
## Year1997          0.13522    0.09027    1.50  0.134
## Year1998          0.03471    0.08766    0.40  0.692
## Year1999         -0.04968    0.08877   -0.56  0.576
```

```

## Year2000          0.02291    0.09278    0.25    0.805
## Year2001         -0.00457    0.08751   -0.05    0.958
## Year2002          0.08541    0.09051    0.94    0.345
## Year2003          0.06955    0.08996    0.77    0.440
## Year2004          0.14665    0.08786    1.67    0.095 .
## Year2005          0.08794    0.08749    1.01    0.315
## Year2006         -0.08184    0.08626   -0.95    0.343
## Year2007          0.00320    0.08304    0.04    0.969
## Year2008         -0.03844    0.08275   -0.46    0.642
## Year2009         -0.04327    0.08299   -0.52    0.602
## Year2010         -0.10731    0.08240   -1.30    0.193
## Year2011          0.02931    0.08250    0.36    0.722
## Year2012         -0.05754    0.08699   -0.66    0.508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0366
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 214 weights are ~= 1. The remaining 2588 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0033 0.8720 0.9530 0.9080 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.57e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.198 1          1.094
## LastAuthorFemale 1.189 1          1.090
## Year              1.043 16          1.001

```

## Residuals from first and last author



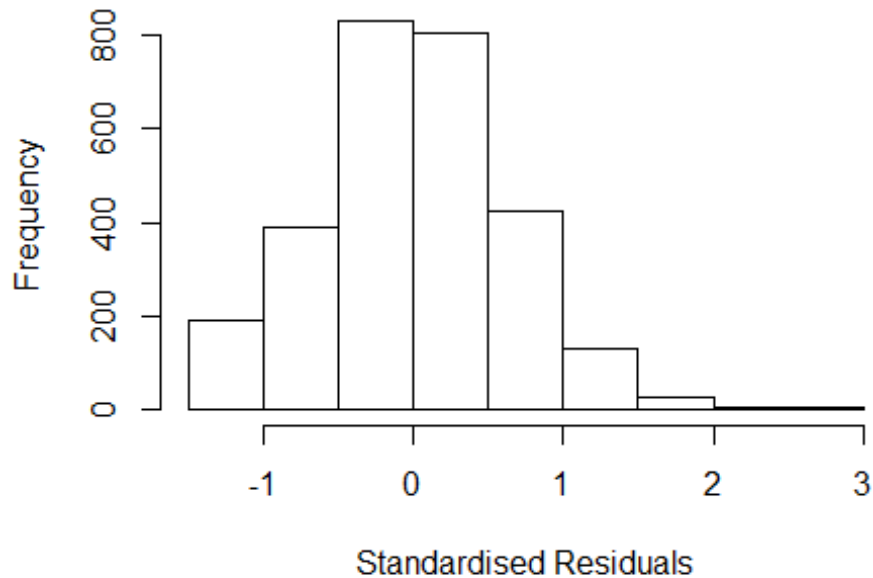
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 826   0032136153 3.657 1998    1702     3    2.533
## 1640 0001224048 3.749 2001    1702     4    2.640
## 1813 0035363218 3.647 2001    1702     5    2.538
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.35815 -0.41539 -0.00733  0.41832  2.64046
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08665    0.07387   14.71  <2e-16 ***
## FirstAuthorFemale1 0.03671    0.03240    1.13   0.257
## LastAuthorFemale1 0.07009    0.03448    2.03   0.042 *
## Year1997         0.16470    0.09177    1.79   0.073 .
## Year1998         0.03776    0.08989    0.42   0.674
## Year1999        -0.01413    0.09102   -0.16   0.877
## Year2000         0.05601    0.09378    0.60   0.550
## Year2001         0.02189    0.08963    0.24   0.807
## Year2002         0.12077    0.09193    1.31   0.189
## Year2003         0.11891    0.09170    1.30   0.195
```

```

## Year2004          0.18628    0.08963    2.08    0.038 *
## Year2005          0.13734    0.08973    1.53    0.126
## Year2006         -0.03400    0.08760   -0.39    0.698
## Year2007          0.02658    0.08492    0.31    0.754
## Year2008          0.01974    0.08440    0.23    0.815
## Year2009         -0.01411    0.08459   -0.17    0.868
## Year2010         -0.06681    0.08525   -0.78    0.433
## Year2011          0.06769    0.08428    0.80    0.422
## Year2012          0.00713    0.08926    0.08    0.936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.0085
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 240 weights are ~= 1. The remaining 2562 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0273 0.8670 0.9510 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1      1.016
## Year              1.032 16      1.001

```

## Residuals from first author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 826   0032136153 3.657 1998    1702     3    2.533
## 1640 0001224048 3.749 2001    1702     4    2.640
## 1813 0035363218 3.647 2001    1702     5    2.538
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34298 -0.41544 -0.00551  0.41822  2.63378
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09268    0.07411   14.74  <2e-16 ***
## FirstAuthorFemale1 0.06387    0.03000    2.13   0.033 *
## Year1997        0.16211    0.09197    1.76   0.078 .
## Year1998        0.03668    0.08997    0.41   0.684
## Year1999       -0.01460    0.09127   -0.16   0.873
## Year2000        0.05778    0.09401    0.61   0.539
## Year2001        0.02255    0.08991    0.25   0.802
## Year2002        0.12275    0.09209    1.33   0.183
## Year2003        0.11996    0.09204    1.30   0.193
## Year2004        0.18643    0.08995    2.07   0.038 *
```

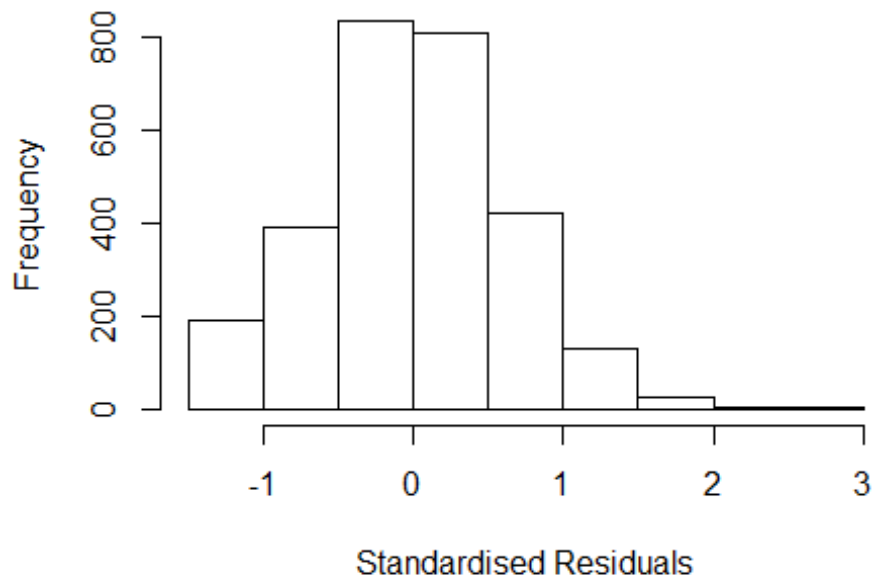


```

## Year2005          0.13625    0.08991    1.52    0.130
## Year2006          -0.03608    0.08783   -0.41    0.681
## Year2007           0.02743    0.08521    0.32    0.748
## Year2008           0.02137    0.08458    0.25    0.801
## Year2009          -0.01207    0.08485   -0.14    0.887
## Year2010          -0.06567    0.08546   -0.77    0.442
## Year2011           0.06753    0.08449    0.80    0.424
## Year2012           0.00733    0.08953    0.08    0.935
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0135, Adjusted R-squared:  0.00746
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 232 weights are ~= 1. The remaining 2570 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0288 0.8660 0.9510 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1          1.012
## Year            1.024 16          1.001

```

## Residuals from last author



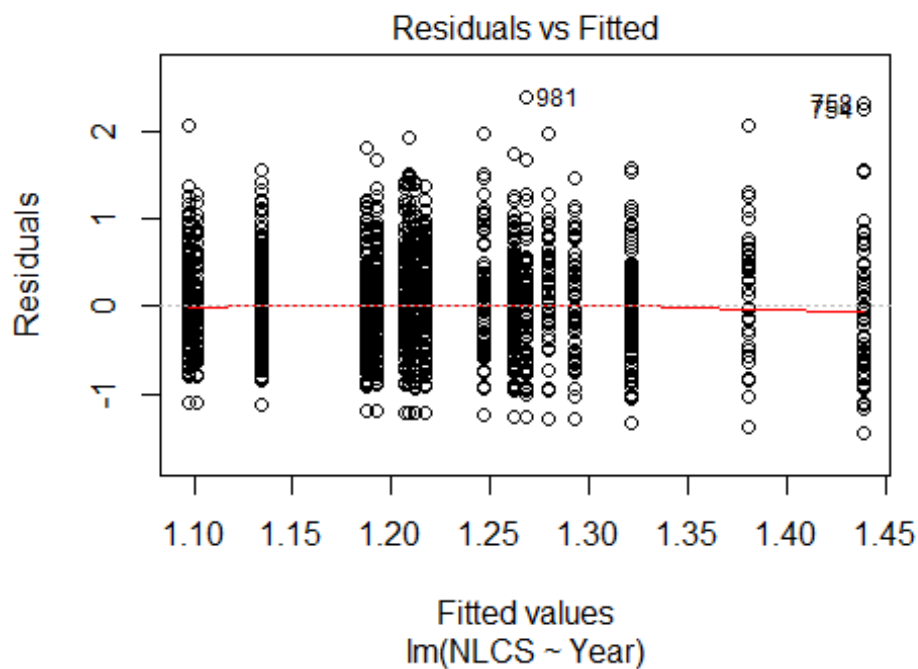
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 826   0032136153 3.657 1998    1702     3    2.533
## 1640 0001224048 3.749 2001    1702     4    2.640
## 1813 0035363218 3.647 2001    1702     5    2.538
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34137 -0.41764 -0.00964  0.41572  2.63567
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.09042    0.07382   14.77  <2e-16 ***
## LastAuthorFemale1 0.08631    0.03195    2.70  0.0069 **
## Year1997        0.16464    0.09188    1.79  0.0733 .
## Year1998        0.03682    0.08997    0.41  0.6824
## Year1999       -0.01551    0.09103   -0.17  0.8647
## Year2000        0.05609    0.09383    0.60  0.5501
## Year2001        0.02291    0.08960    0.26  0.7982
## Year2002        0.12064    0.09194    1.31  0.1896
## Year2003        0.11738    0.09168    1.28  0.2006
## Year2004        0.18870    0.08967    2.10  0.0354 *
```

```

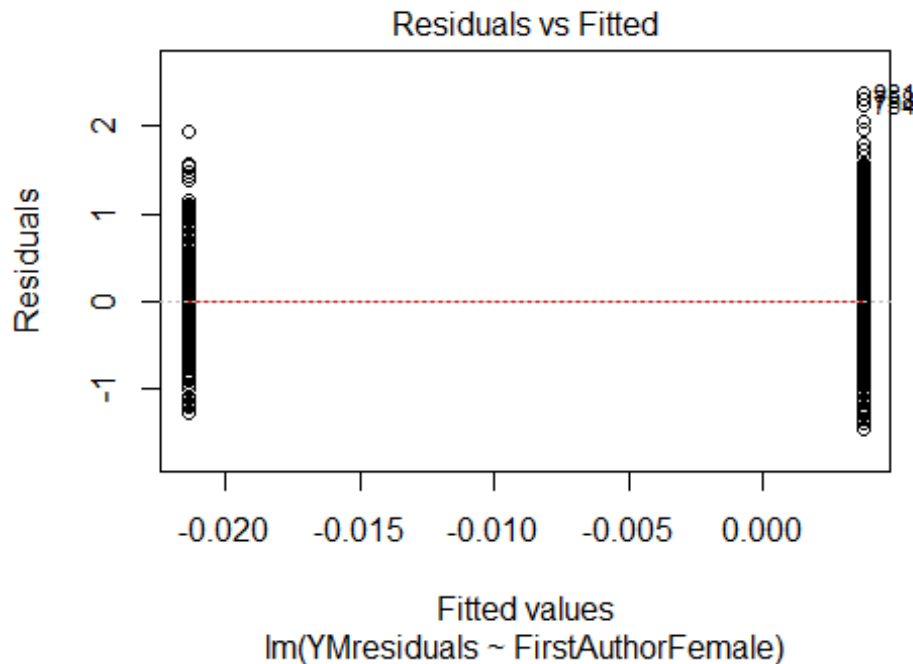
## Year2005      0.13967      0.08980      1.56      0.1200
## Year2006     -0.03039      0.08762     -0.35      0.7288
## Year2007      0.02724      0.08498      0.32      0.7486
## Year2008      0.02150      0.08443      0.25      0.7990
## Year2009     -0.01370      0.08461     -0.16      0.8714
## Year2010     -0.06556      0.08532     -0.77      0.4423
## Year2011      0.07128      0.08422      0.85      0.3974
## Year2012      0.00819      0.08929      0.09      0.9269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0144, Adjusted R-squared:  0.00837
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 245 weights are ~= 1. The remaining 2557 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0288 0.8670 0.9510 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2802"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1703"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 148 127 166 128 147 190 176 190 179 213 218 277 261 232 228
## 2011 2012
## 220 239
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 47 56 75 60 59 71 75 116 96 128 136 186 168 150 148
## 2011 2012
## 149 180
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 51 70 55 53 65 66 112 87 107 125 155 147 133 131
## 2011 2012
## 131 154
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 50, df = 16, p-value = 2e-05
```

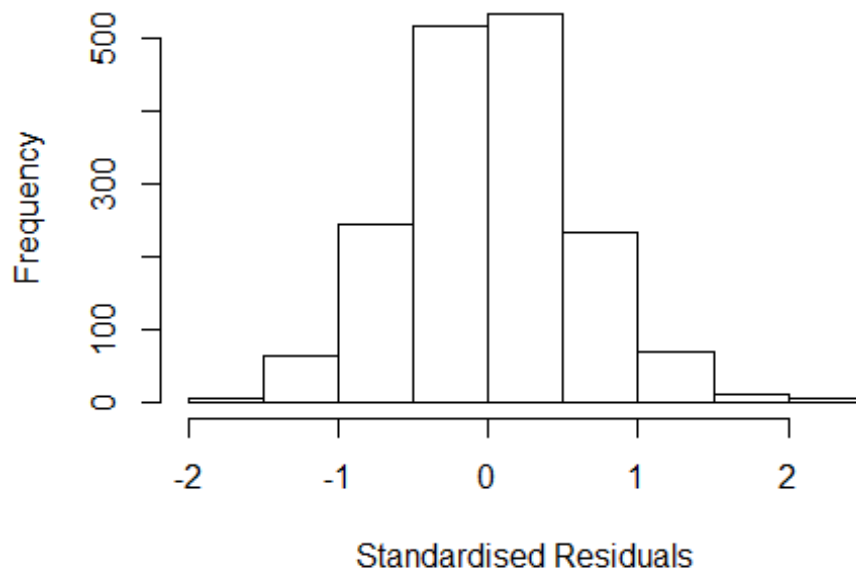


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.071, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 2.26557996667244"
## [1] "Male first author team size 2018 geometric mean: 2.87123205481943"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 830, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.3091289524044"
## [1] "Male last author team size 2018 geometric mean: 2.87788284029346"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 940, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.130 1      1.063
## LastAuthorFemale  1.085 1      1.042
## UniqueAuthors    1.266 4      1.030
## Year              1.292 16     1.008
```

## Residuals from first and last author and team size



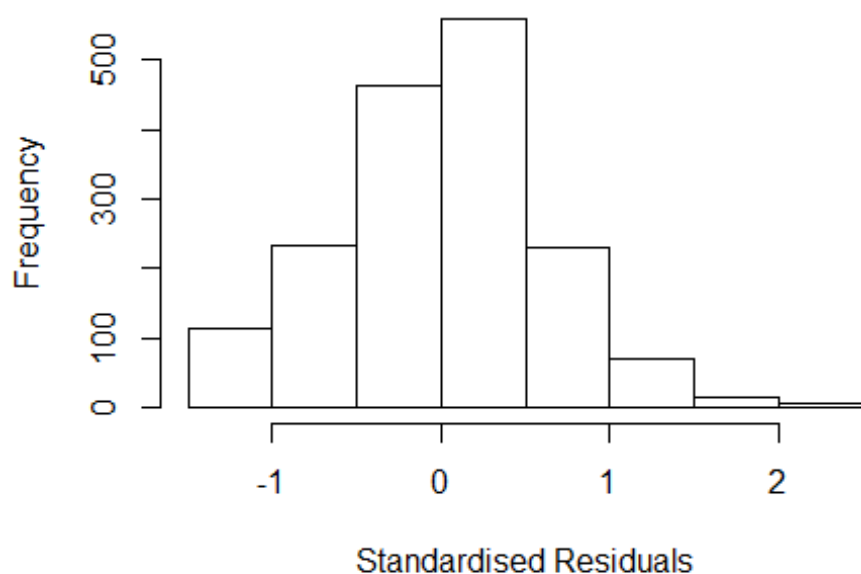
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6776 -0.3779  0.0112  0.3789  2.2126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2217    0.1660   7.36 2.9e-13 ***
## FirstAuthorFemale1 -0.0504    0.0408  -1.24  0.217
## LastAuthorFemale1 -0.0575    0.0441  -1.30  0.193
## UniqueAuthors2     0.2818    0.0384   7.35 3.1e-13 ***
## UniqueAuthors3     0.3554    0.0430   8.27 2.8e-16 ***
## UniqueAuthors4     0.3898    0.0555   7.02 3.3e-12 ***
## UniqueAuthors5     0.5986    0.0566  10.57 < 2e-16 ***
## Year1997          -0.0749    0.1940  -0.39  0.700
## Year1998          -0.1314    0.1930  -0.68  0.496
## Year1999          -0.0667    0.1929  -0.35  0.730
```

```

## Year2000          -0.0551      0.1993   -0.28    0.782
## Year2001          -0.1427      0.1872   -0.76    0.446
## Year2002          -0.3407      0.1820   -1.87    0.061 .
## Year2003          -0.1463      0.1752   -0.84    0.404
## Year2004          -0.2287      0.1775   -1.29    0.198
## Year2005          -0.2656      0.1734   -1.53    0.126
## Year2006          -0.2401      0.1731   -1.39    0.166
## Year2007          -0.2850      0.1731   -1.65    0.100 .
## Year2008          -0.3319      0.1729   -1.92    0.055 .
## Year2009          -0.2578      0.1726   -1.49    0.135
## Year2010          -0.2322      0.1731   -1.34    0.180
## Year2011          -0.3841      0.1720   -2.23    0.026 *
## Year2012          -0.3161      0.1731   -1.83    0.068 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0944
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 148 weights are ~= 1. The remaining 1538 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.085  0.859  0.951   0.898   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.93e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.068 1          1.034
## LastAuthorFemale  1.071 1          1.035
## Year              1.076 16          1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3726 -0.3927 0.0278 0.3838 2.3927
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36929 0.14684 9.33 <2e-16 ***
## FirstAuthorFemale1 0.00302 0.04303 0.07 0.944
## LastAuthorFemale1 -0.06017 0.04761 -1.26 0.206
## Year1997 -0.07394 0.17939 -0.41 0.680
## Year1998 -0.14541 0.17725 -0.82 0.412
## Year1999 -0.03192 0.18204 -0.18 0.861
## Year2000 0.00328 0.18266 0.02 0.986
## Year2001 -0.11503 0.17317 -0.66 0.507
## Year2002 -0.28462 0.16515 -1.72 0.085 .
## Year2003 -0.05066 0.15599 -0.32 0.745
## Year2004 -0.15908 0.16079 -0.99 0.323
## Year2005 -0.16772 0.15535 -1.08 0.280
```

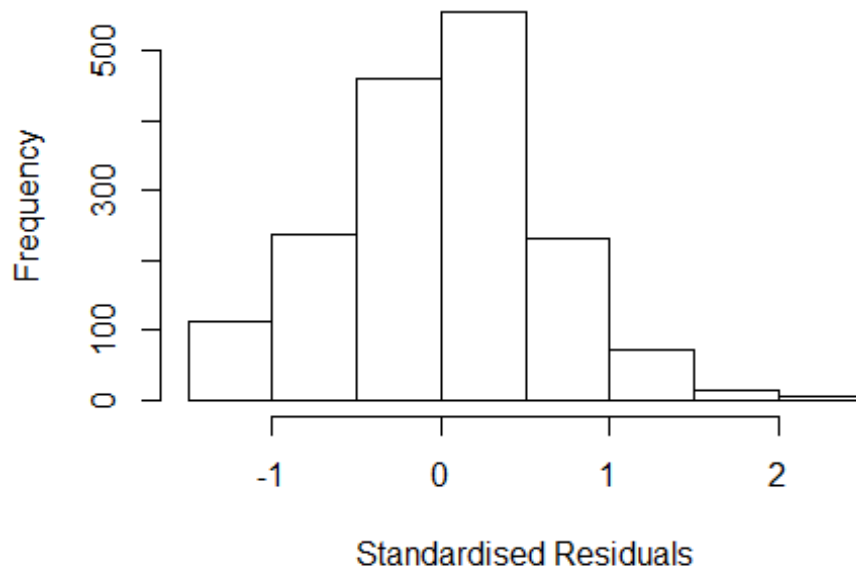


```

## Year2006      -0.17233      0.15566      -1.11      0.268
## Year2007      -0.20094      0.15474      -1.30      0.194
## Year2008      -0.21559      0.15482      -1.39      0.164
## Year2009      -0.14018      0.15475      -0.91      0.365
## Year2010      -0.09566      0.15477      -0.62      0.537
## Year2011      -0.28048      0.15438      -1.82      0.069 .
## Year2012      -0.17140      0.15516      -1.10      0.269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0157, Adjusted R-squared:  0.00511
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1542 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0603 0.8570 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.93e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1      1.019
## Year      1.037 16      1.001

```

## Residuals from first author



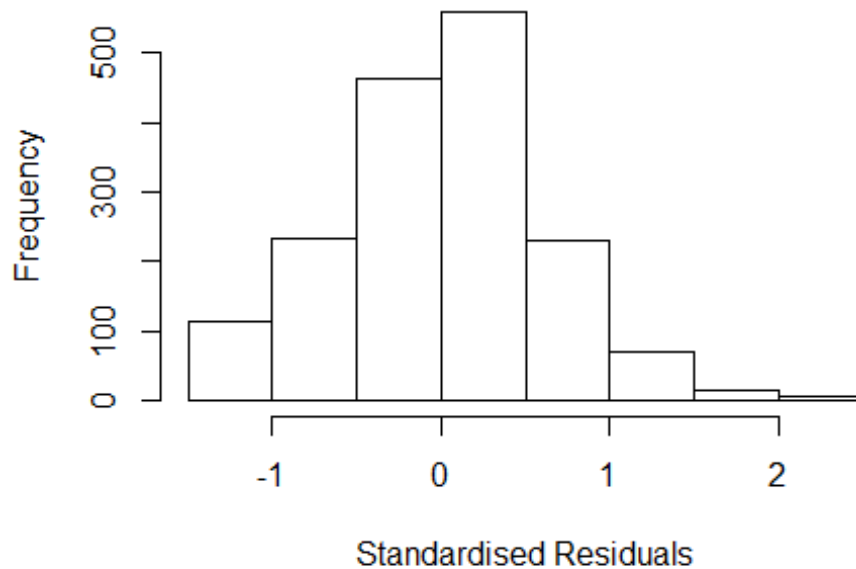
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3665 -0.3938 0.0285 0.3893 2.3959
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.365699 0.146769 9.31 <2e-16 ***
## FirstAuthorFemale1 -0.011863 0.043149 -0.27 0.783
## Year1997 -0.077883 0.179849 -0.43 0.665
## Year1998 -0.147432 0.177342 -0.83 0.406
## Year1999 -0.035835 0.181570 -0.20 0.844
## Year2000 0.000808 0.182379 0.00 0.996
## Year2001 -0.114636 0.173352 -0.66 0.509
## Year2002 -0.286376 0.165190 -1.73 0.083 .
## Year2003 -0.050520 0.155874 -0.32 0.746
## Year2004 -0.162448 0.160787 -1.01 0.312
## Year2005 -0.166778 0.155363 -1.07 0.283
## Year2006 -0.176289 0.155688 -1.13 0.258
```

```

## Year2007          -0.204115    0.154771    -1.32    0.187
## Year2008          -0.215469    0.154794    -1.39    0.164
## Year2009          -0.143180    0.154858    -0.92    0.355
## Year2010          -0.098773    0.154911    -0.64    0.524
## Year2011          -0.279152    0.154454    -1.81    0.071 .
## Year2012          -0.172411    0.155153    -1.11    0.267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.588
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.00476
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0589 0.8580 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.93e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.04 1          1.020
## Year              1.04 16          1.001

```

## Residuals from last author



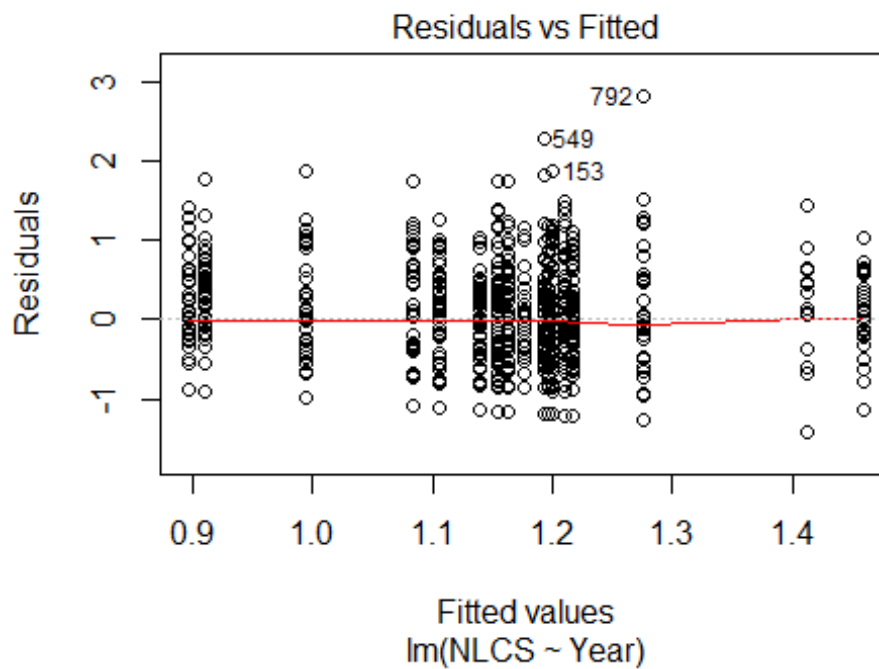
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3727 -0.3937 0.0283 0.3834 2.3925
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36962 0.14673 9.33 <2e-16 ***
## LastAuthorFemale1 -0.05927 0.04710 -1.26 0.208
## Year1997 -0.07393 0.17949 -0.41 0.680
## Year1998 -0.14539 0.17733 -0.82 0.412
## Year1999 -0.03191 0.18212 -0.18 0.861
## Year2000 0.00304 0.18271 0.02 0.987
## Year2001 -0.11517 0.17322 -0.66 0.506
## Year2002 -0.28475 0.16518 -1.72 0.085 .
## Year2003 -0.05068 0.15606 -0.32 0.745
## Year2004 -0.15931 0.16074 -0.99 0.322
## Year2005 -0.16761 0.15543 -1.08 0.281
## Year2006 -0.17206 0.15569 -1.11 0.269
```

```

## Year2007          -0.20100      0.15478    -1.30      0.194
## Year2008          -0.21563      0.15488    -1.39      0.164
## Year2009          -0.14000      0.15486    -0.90      0.366
## Year2010          -0.09570      0.15482    -0.62      0.537
## Year2011          -0.28049      0.15447    -1.82      0.070 .
## Year2012          -0.17124      0.15522    -1.10      0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0157, Adjusted R-squared:  0.00571
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1543 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0597 0.8570 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.93e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1686"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1704"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   69   90   98  101   78  104   43   54   84   76  101   85   76  102   90
## 2011 2012
##  113  102
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   37   50   41   54   29   54   18   28   39   54   59   48   47   58   53
## 2011 2012

```

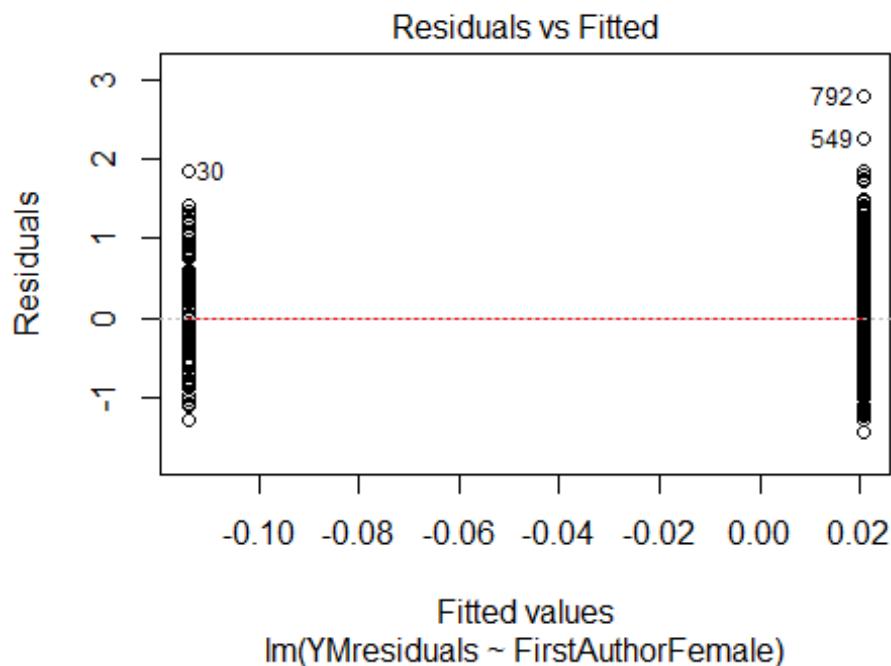
```
## 68 65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33 47 39 51 28 53 17 26 36 46 53 39 40 53 46
## 2011 2012
## 59 56
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.54, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 2.56377087722506"
## [1] "Male first author team size 2018 geometric mean: 2.96097962099618"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

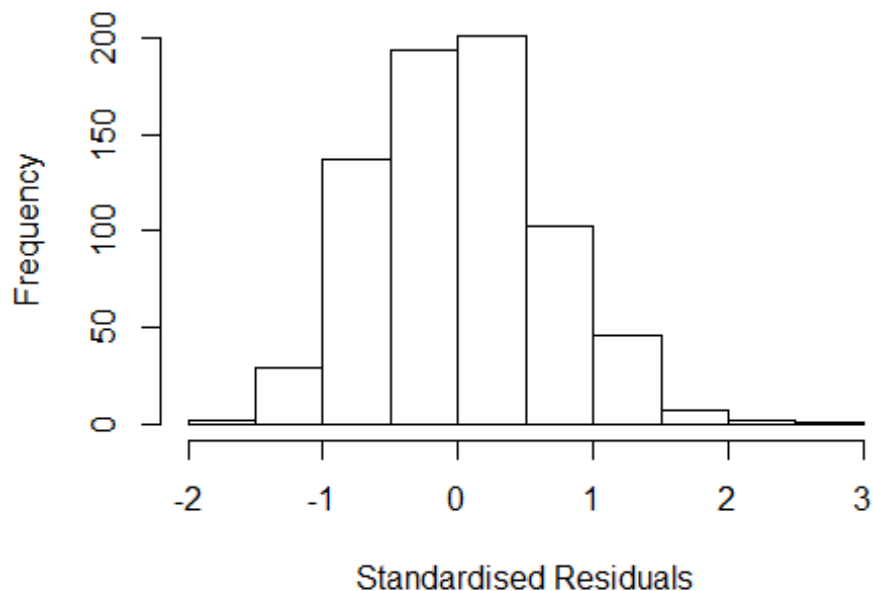
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 240, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.85636609131688"
## [1] "Male last author team size 2018 geometric mean: 3.14804179374093"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.378 1      1.174
## LastAuthorFemale  1.400 1      1.183
## UniqueAuthors    1.632 4      1.063
## Year              2.001 16     1.022
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 4344675264 4.089 2004      1704      3      2.719
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.53846 -0.46019 -0.00916  0.44051  2.71914
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90511    0.17661   5.13 3.9e-07 ***
## FirstAuthorFemale1 -0.13756    0.08355  -1.65   0.10
## LastAuthorFemale1  0.00628    0.09021   0.07   0.94
## UniqueAuthors2    0.30094    0.06818   4.41 1.2e-05 ***
## UniqueAuthors3    0.45827    0.06790   6.75 3.1e-11 ***
## UniqueAuthors4    0.47728    0.09517   5.01 6.7e-07 ***
## UniqueAuthors5    0.59883    0.11342   5.28 1.7e-07 ***
## Year1997          0.03452    0.22003   0.16   0.88
## Year1998         -0.20062    0.20683  -0.97   0.33
## Year1999         -0.00512    0.20055  -0.03   0.98
```

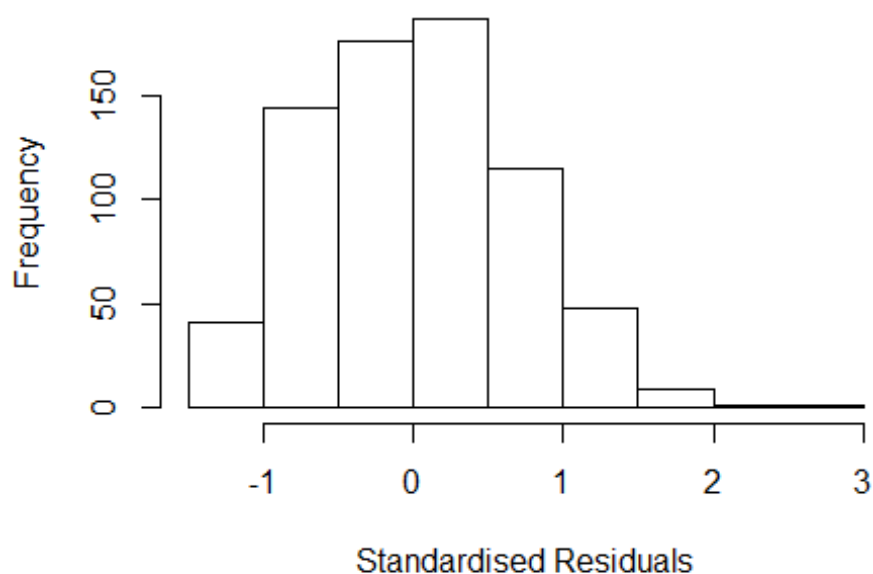


```

## Year2000      -0.03723    0.20158   -0.18    0.85
## Year2001      0.02659    0.19397    0.14    0.89
## Year2002      0.28959    0.27727    1.04    0.30
## Year2003      0.24273    0.20525    1.18    0.24
## Year2004      0.00648    0.23132    0.03    0.98
## Year2005     -0.00597    0.20686   -0.03    0.98
## Year2006     -0.00229    0.20243   -0.01    0.99
## Year2007     -0.23960    0.20715   -1.16    0.25
## Year2008      0.03378    0.20286    0.17    0.87
## Year2009     -0.07064    0.19861   -0.36    0.72
## Year2010     -0.04421    0.19123   -0.23    0.82
## Year2011      0.08916    0.19306    0.46    0.64
## Year2012     -0.31363    0.19353   -1.62    0.11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.655
## Multiple R-squared:  0.138, Adjusted R-squared:  0.111
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 664 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0461 0.8720 0.9490 0.9080 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.39e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.275 1      1.129
## LastAuthorFemale  1.288 1      1.135
## Year              1.236 16      1.007

```

## Residuals from first and last author

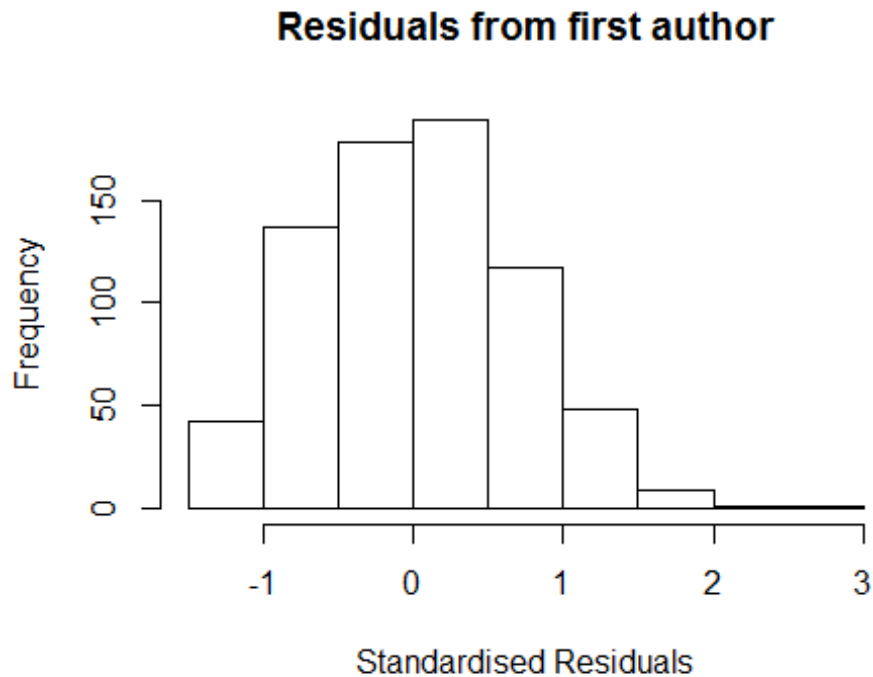


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 4344675264 4.089 2004      1704      3      2.833
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4954 -0.5024  0.0024  0.4871  2.8328
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1072     0.1575   7.03 5e-12 ***
## FirstAuthorFemale1 -0.1283     0.0866  -1.48  0.139
## LastAuthorFemale1 -0.0631     0.0933  -0.68  0.499
## Year1997          0.0704     0.2011   0.35  0.727
## Year1998         -0.2220     0.1977  -1.12  0.262
## Year1999          0.0646     0.1871   0.35  0.730
## Year2000          0.0313     0.1860   0.17  0.867
## Year2001          0.0715     0.1797   0.40  0.691
## Year2002          0.3882     0.2907   1.34  0.182
## Year2003          0.3434     0.1928   1.78  0.075 .
## Year2004          0.1490     0.2179   0.68  0.494
## Year2005          0.0420     0.1959   0.21  0.830
```

```

## Year2006          0.0145      0.1898      0.08      0.939
## Year2007         -0.1737      0.1936     -0.90      0.370
## Year2008          0.1190      0.1871      0.64      0.525
## Year2009          0.0448      0.1897      0.24      0.814
## Year2010          0.0263      0.1768      0.15      0.882
## Year2011          0.1730      0.1827      0.95      0.344
## Year2012         -0.2543      0.1882     -1.35      0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.711
## Multiple R-squared:  0.0505, Adjusted R-squared:  0.0262
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 662 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0772 0.8750 0.9490 0.9140 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.39e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.118 1      1.057
## Year              1.118 16      1.003

```

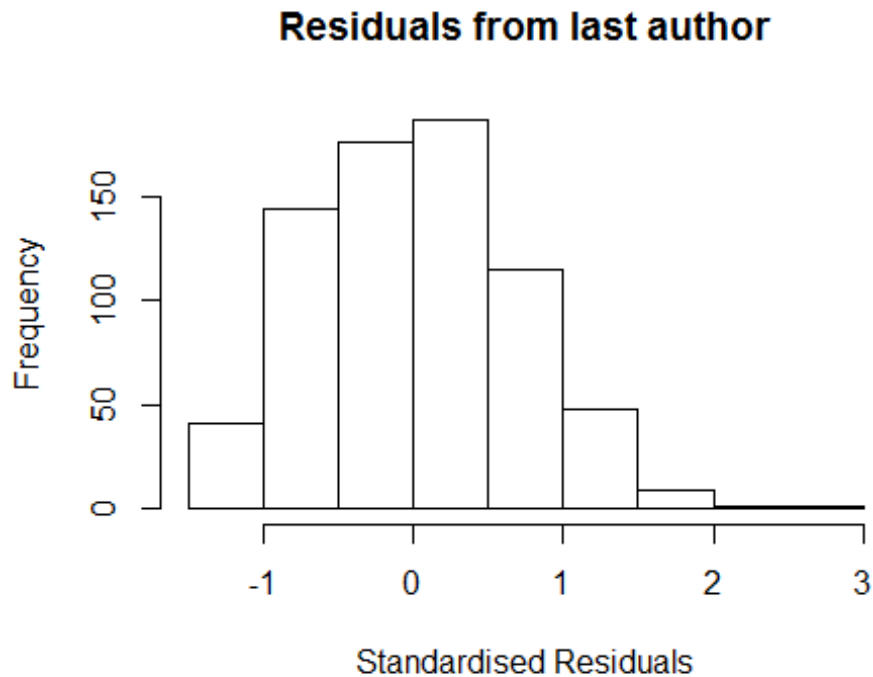


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 4344675264 4.089 2004      1704      3      2.833
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4947 -0.4985  0.0115  0.4870  2.8355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1062     0.1566   7.06 3.9e-12 ***
## FirstAuthorFemale1 -0.1541     0.0823  -1.87  0.062 .
## Year1997          0.0676     0.2008   0.34  0.736
## Year1998         -0.2256     0.1965  -1.15  0.251
## Year1999          0.0614     0.1857   0.33  0.741
## Year2000          0.0308     0.1847   0.17  0.868
## Year2001          0.0711     0.1788   0.40  0.691
## Year2002          0.3885     0.2911   1.33  0.182
## Year2003          0.3474     0.1919   1.81  0.071 .
## Year2004          0.1473     0.2179   0.68  0.499
## Year2005          0.0339     0.1943   0.17  0.861
## Year2006          0.0123     0.1888   0.06  0.948
```

```

## Year2007          -0.1783      0.1923    -0.93     0.354
## Year2008           0.1173      0.1861      0.63     0.529
## Year2009           0.0325      0.1863      0.17     0.862
## Year2010           0.0261      0.1756      0.15     0.882
## Year2011           0.1685      0.1809      0.93     0.352
## Year2012          -0.2627      0.1865     -1.41     0.159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0496, Adjusted R-squared:  0.0267
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 667 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0771 0.8770 0.9490 0.9150 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.39e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.129 1      1.062
## Year              1.129 16      1.004

```



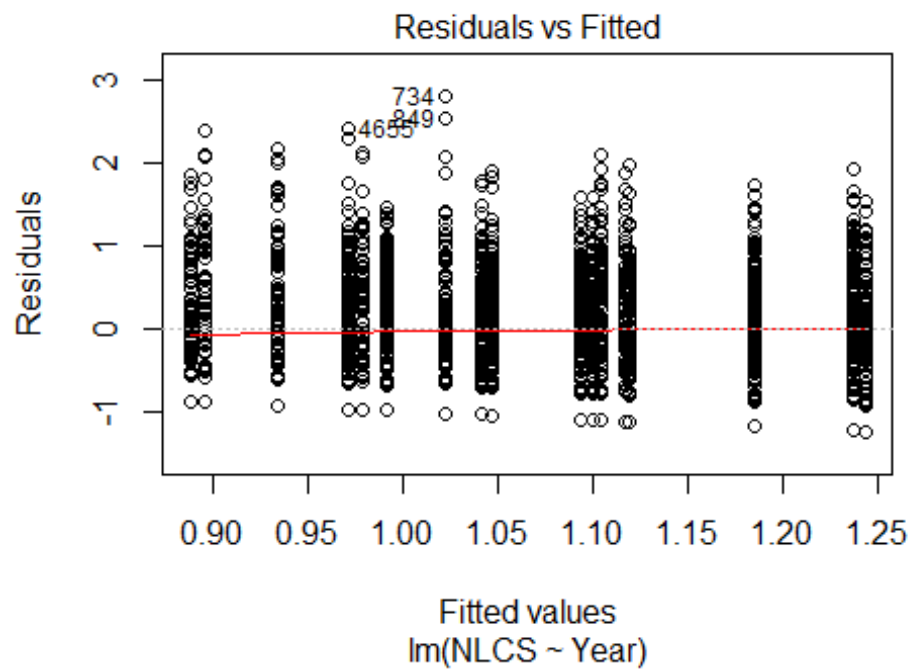
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 4344675264 4.089 2004      1704      3      2.833
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49405 -0.48910  0.00055  0.48178  2.85981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10535    0.15562   7.10 3e-12 ***
## LastAuthorFemale1 -0.12143    0.08966  -1.35  0.176
## Year1997         0.06299    0.19758   0.32  0.750
## Year1998        -0.22429    0.19599  -1.14  0.253
## Year1999         0.06666    0.18512   0.36  0.719
## Year2000         0.02711    0.18448   0.15  0.883
## Year2001         0.06582    0.17708   0.37  0.710
## Year2002         0.38870    0.29128   1.33  0.182
## Year2003         0.32093    0.18935   1.69  0.091 .
## Year2004         0.12384    0.21616   0.57  0.567
## Year2005         0.04153    0.19492   0.21  0.831
## Year2006        -0.00387    0.18758  -0.02  0.984
```

```

## Year2007          -0.18831    0.19089   -0.99    0.324
## Year2008           0.10478    0.18358    0.57    0.568
## Year2009           0.03259    0.18698    0.17    0.862
## Year2010           0.01136    0.17395    0.07    0.948
## Year2011           0.16155    0.18108    0.89    0.373
## Year2012          -0.26121    0.18710   -1.40    0.163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.714
## Multiple R-squared:  0.0473, Adjusted R-squared:  0.0243
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 662 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0719 0.8770 0.9490 0.9150 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.39e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 722"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1705"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 254 294 256 245 287 281 206 221 168 237 289 297 257 261 284
## 2011 2012
## 260 229
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 87 100 106 108 133 97 128 119 97 163 153 203 155 173 190
## 2011 2012

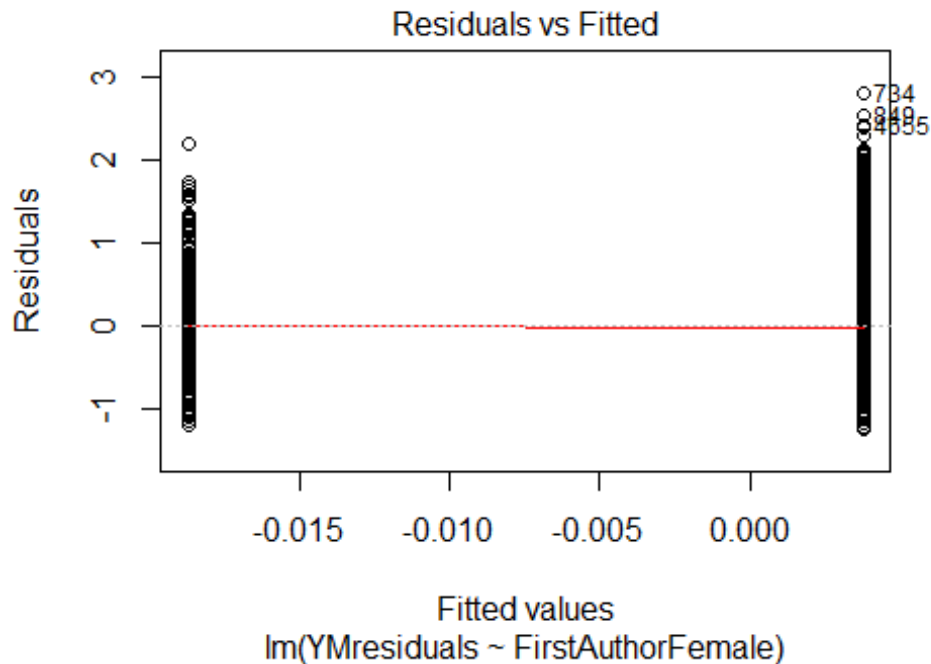
```

```
## 175 164
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 79 94 101 99 125 86 116 106 85 144 138 178 134 159 163
## 2011 2012
## 152 139
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 52, df = 16, p-value = 1e-05
```



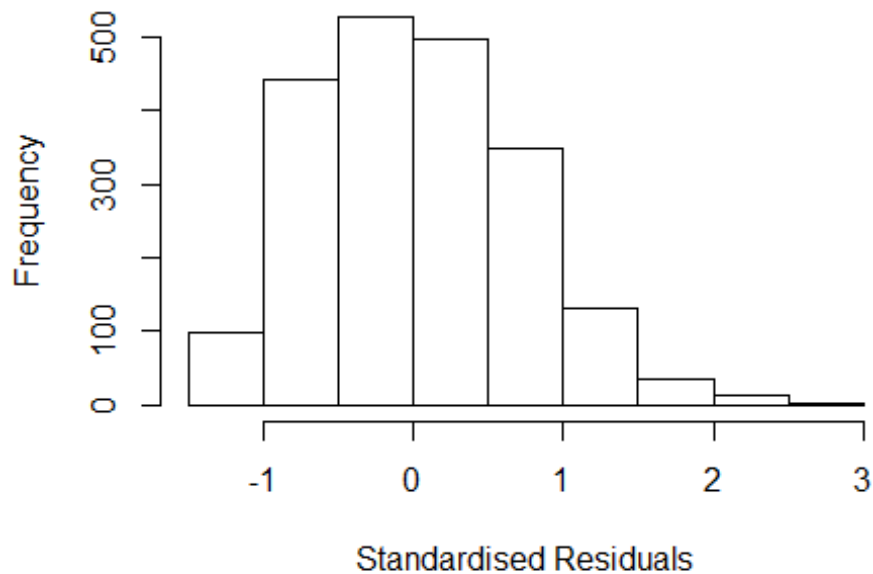
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```





```
## [1] "Female first author team size 2018 geometric mean: 2.31773180700604"
## [1] "Male first author team size 2018 geometric mean: 2.41800038677814"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.0891372726288"
## [1] "Male last author team size 2018 geometric mean: 2.45582775685033"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 390, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.254 1      1.120
## LastAuthorFemale  1.278 1      1.131
## UniqueAuthors     1.214 4      1.025
## Year               1.231 16     1.007
```

## Residuals from first and last author and team size



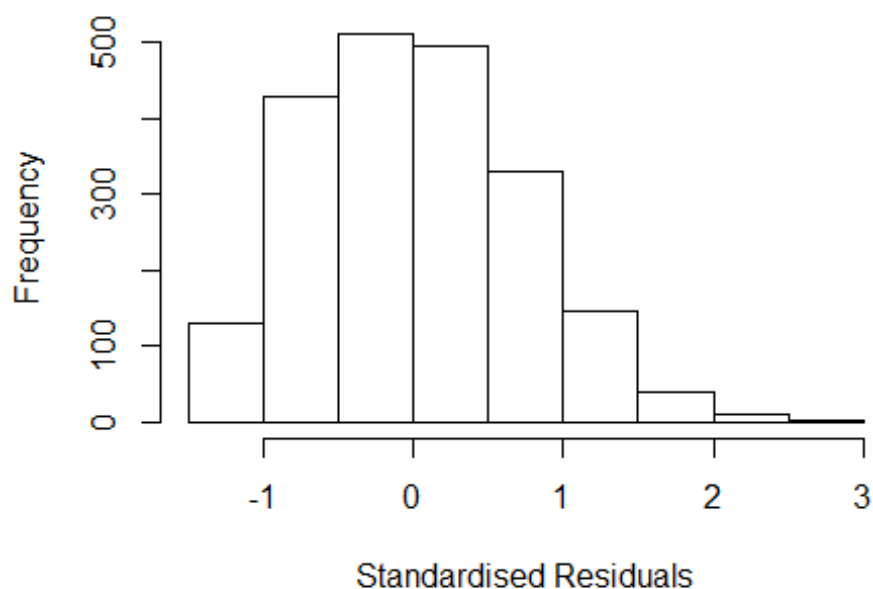
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 734   0032070303 3.831 1998    1705      2    2.803
## 849   0031643297 3.557 1998    1705      4    2.746
## 4575  76449099878 3.269 2010    1705      3    2.551
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3769 -0.5160 -0.0222  0.5083  2.8032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6741    0.1066   6.32 3.1e-10 ***
## FirstAuthorFemale1  0.0153    0.0470   0.32  0.7453
## LastAuthorFemale1 -0.0543    0.0517  -1.05  0.2939
## UniqueAuthors2    0.2172    0.0402   5.40 7.3e-08 ***
## UniqueAuthors3    0.3613    0.0451   8.02 1.8e-15 ***
## UniqueAuthors4    0.4833    0.0647   7.47 1.2e-13 ***
## UniqueAuthors5    0.4064    0.0664   6.12 1.1e-09 ***
## Year1997         0.2493    0.1429   1.74  0.0813 .
```

```

## Year1998          0.1365      0.1257      1.09      0.2775
## Year1999          0.0294      0.1293      0.23      0.8200
## Year2000          0.0500      0.1319      0.38      0.7048
## Year2001          0.0579      0.1391      0.42      0.6771
## Year2002          0.1943      0.1323      1.47      0.1420
## Year2003          0.2393      0.1239      1.93      0.0535 .
## Year2004          0.3415      0.1298      2.63      0.0086 **
## Year2005          0.2616      0.1177      2.22      0.0264 *
## Year2006          0.1577      0.1195      1.32      0.1873
## Year2007          0.1827      0.1162      1.57      0.1162
## Year2008          0.0942      0.1219      0.77      0.4397
## Year2009          0.2327      0.1189      1.96      0.0504 .
## Year2010          0.0439      0.1183      0.37      0.7106
## Year2011          0.1277      0.1179      1.08      0.2789
## Year2012          0.2632      0.1235      2.13      0.0332 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.722
## Multiple R-squared:  0.0727, Adjusted R-squared:  0.0629
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1951 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0981 0.8760 0.9470 0.9140 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.77e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.258 1 1.122
## LastAuthorFemale 1.252 1 1.119
## Year 1.063 16 1.002

```

## Residuals from first and last author



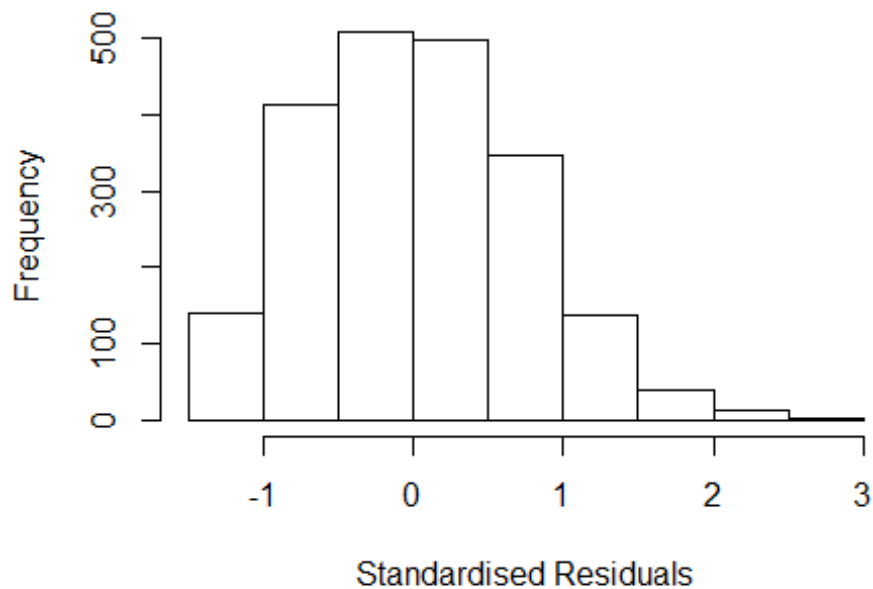
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 182 0000270113 3.290 1996    1705     3    2.573
## 734 0032070303 3.831 1998    1705     2    2.885
## 849 0031643297 3.557 1998    1705     4    2.611
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2185 -0.5304 -0.0281  0.5074  2.8853
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8135    0.1116   7.29 4.4e-13 ***
## FirstAuthorFemale1  0.0254    0.0475   0.54  0.5927
## LastAuthorFemale1 -0.0967    0.0520  -1.86  0.0628 .
## Year1997         0.2486    0.1476   1.68  0.0922 .
## Year1998         0.1322    0.1328   1.00  0.3198
## Year1999         0.0308    0.1383   0.22  0.8236
## Year2000         0.0758    0.1380   0.55  0.5830
## Year2001         0.0762    0.1440   0.53  0.5970
## Year2002         0.2500    0.1373   1.82  0.0688 .
## Year2003         0.2926    0.1305   2.24  0.0251 *
```

```

## Year2004          0.4050      0.1347      3.01      0.0027 **
## Year2005          0.3523      0.1240      2.84      0.0045 **
## Year2006          0.2042      0.1255      1.63      0.1039
## Year2007          0.2562      0.1222      2.10      0.0361 *
## Year2008          0.1488      0.1283      1.16      0.2462
## Year2009          0.2984      0.1242      2.40      0.0163 *
## Year2010          0.1393      0.1227      1.13      0.2565
## Year2011          0.2023      0.1239      1.63      0.1028
## Year2012          0.3802      0.1283      2.96      0.0031 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.749
## Multiple R-squared:  0.023, Adjusted R-squared:  0.0146
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 154 weights are ~= 1. The remaining 1944 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.104  0.876  0.948   0.915   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.77e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```

## Residuals from first author



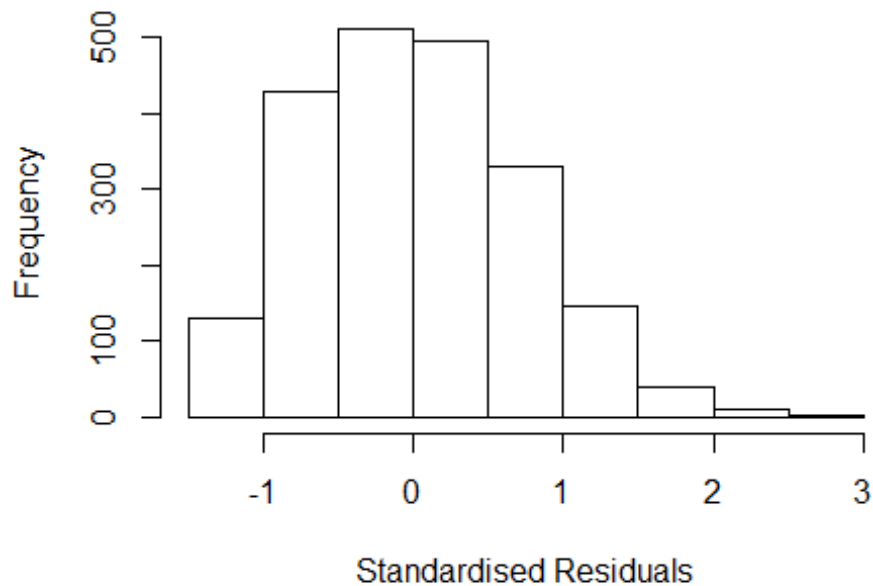
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 182 0000270113 3.290 1996    1705     3    2.573
## 734 0032070303 3.831 1998    1705     2    2.885
## 849 0031643297 3.557 1998    1705     4    2.611
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2076 -0.5366 -0.0254  0.5125  2.8936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8113    0.1116   7.27   5e-13 ***
## FirstAuthorFemale1 -0.0165    0.0438  -0.38   0.7064
## Year1997        0.2440    0.1483   1.65   0.1000 .
## Year1998        0.1261    0.1332   0.95   0.3438
## Year1999        0.0320    0.1382   0.23   0.8168
## Year2000        0.0773    0.1383   0.56   0.5765
## Year2001        0.0708    0.1440   0.49   0.6231
## Year2002        0.2500    0.1372   1.82   0.0687 .
## Year2003        0.2919    0.1308   2.23   0.0258 *
## Year2004        0.3963    0.1344   2.95   0.0032 **
```

```

## Year2005          0.3429      0.1238      2.77      0.0056 **
## Year2006          0.1964      0.1255      1.57      0.1176
## Year2007          0.2523      0.1221      2.07      0.0388 *
## Year2008          0.1461      0.1282      1.14      0.2548
## Year2009          0.2937      0.1239      2.37      0.0178 *
## Year2010          0.1317      0.1226      1.07      0.2827
## Year2011          0.1921      0.1238      1.55      0.1210
## Year2012          0.3721      0.1284      2.90      0.0038 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.75
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.0132
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 154 weights are ~ = 1. The remaining 1944 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.104  0.876   0.949   0.915   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.77e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1      1.016
## Year      1.033 16      1.001

```

## Residuals from last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 182 0000270113 3.290 1996    1705     3    2.573
## 734 0032070303 3.831 1998    1705     2    2.885
## 849 0031643297 3.557 1998    1705     4    2.611
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2205 -0.5319 -0.0272  0.5056  2.8832
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8154    0.1113    7.32 3.4e-13 ***
## LastAuthorFemale1 -0.0843    0.0471   -1.79  0.0737 .
## Year1997        0.2475    0.1474    1.68  0.0932 .
## Year1998        0.1325    0.1328    1.00  0.3186
## Year1999        0.0318    0.1382    0.23  0.8181
## Year2000        0.0753    0.1379    0.55  0.5850
## Year2001        0.0765    0.1440    0.53  0.5952
## Year2002        0.2510    0.1373    1.83  0.0676 .
## Year2003        0.2922    0.1305    2.24  0.0252 *
## Year2004        0.4051    0.1346    3.01  0.0027 **
```

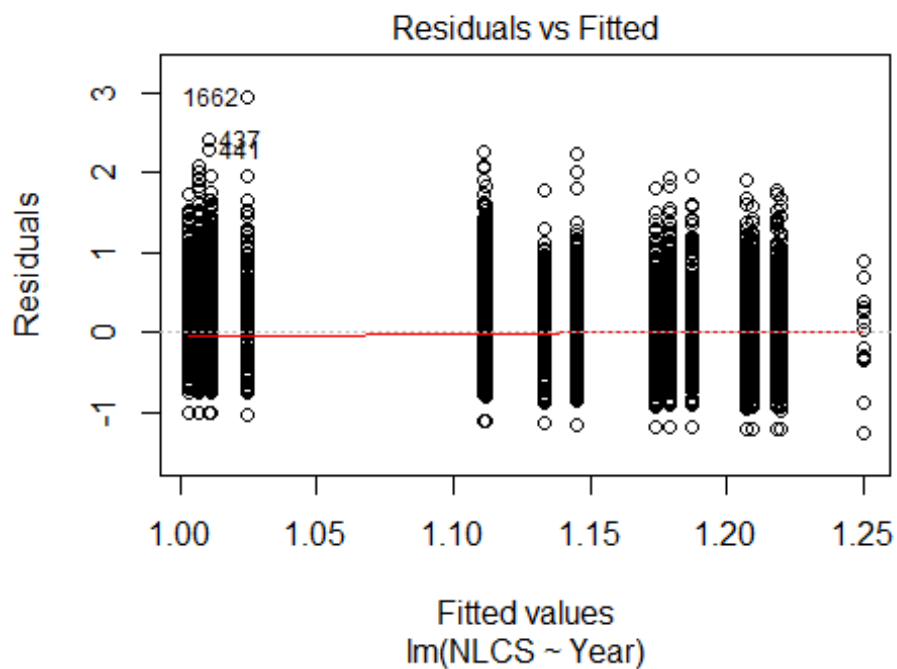


```

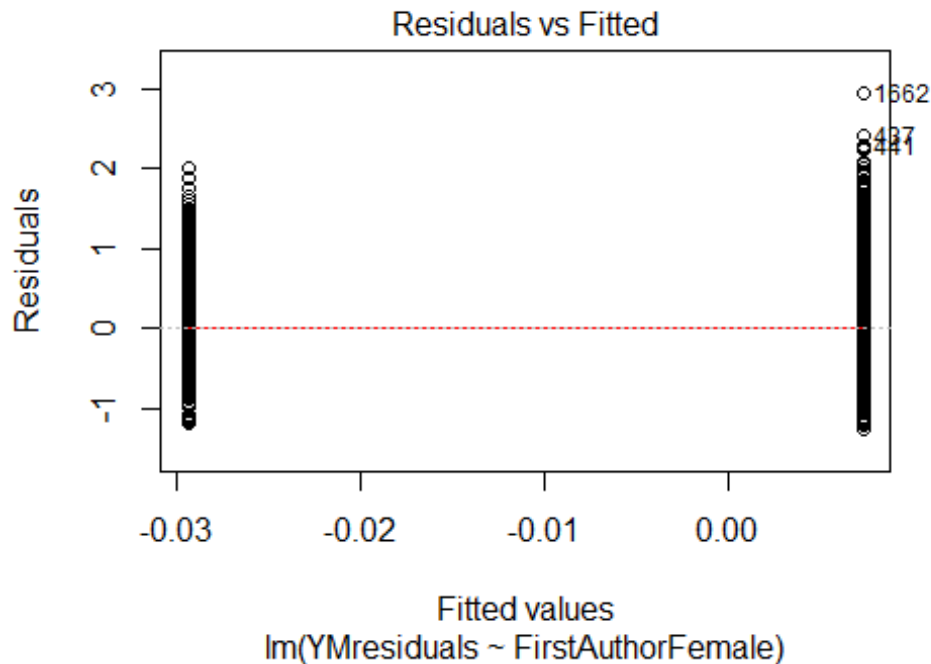
## Year2005          0.3517      0.1240      2.84      0.0046 **
## Year2006          0.2046      0.1255      1.63      0.1032
## Year2007          0.2578      0.1222      2.11      0.0350 *
## Year2008          0.1495      0.1283      1.17      0.2441
## Year2009          0.2995      0.1241      2.41      0.0159 *
## Year2010          0.1409      0.1227      1.15      0.2508
## Year2011          0.2039      0.1239      1.65      0.0999 .
## Year2012          0.3814      0.1283      2.97      0.0030 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.749
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.0149
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1949 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.105  0.875   0.948   0.915   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.77e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2098"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1706"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   701   36  693  569  712  896  734  552  618  651  716  820  872  957  937
## 2011 2012
##   957 1024
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 230 15 291 238 287 376 363 284 313 346 405 452 476 557 558
## 2011 2012
## 563 624
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 208 15 274 223 261 347 317 256 278 296 346 373 403 482 471
## 2011 2012
## 476 552
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 110, df = 16, p-value = 1e-15
```

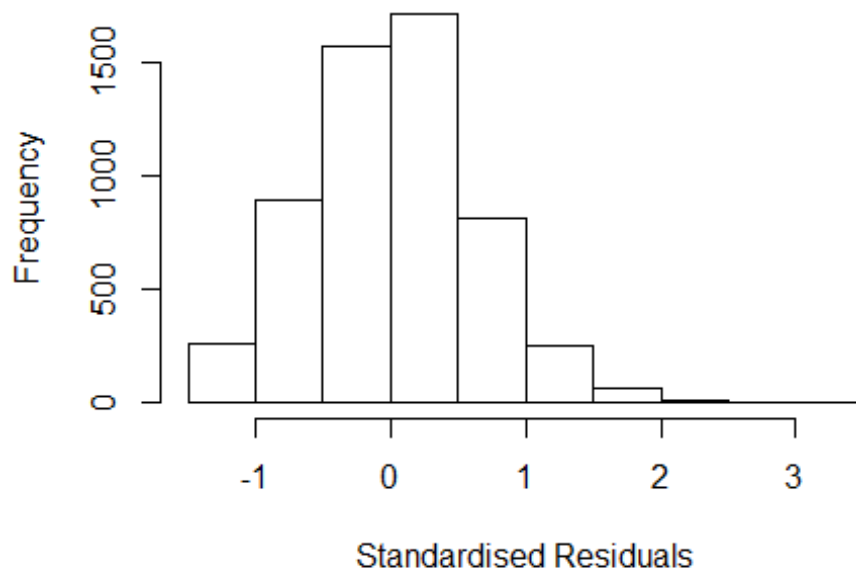


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.16113533503465"
## [1] "Male first author team size 2018 geometric mean: 2.51717541243665"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6600, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.10759614369777"
## [1] "Male last author team size 2018 geometric mean: 2.54273165024073"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6900, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.127 1      1.062
## LastAuthorFemale  1.131 1      1.064
## UniqueAuthors     1.139 4      1.016
## Year              1.157 16     1.005
```

## Residuals from first and last author and team size



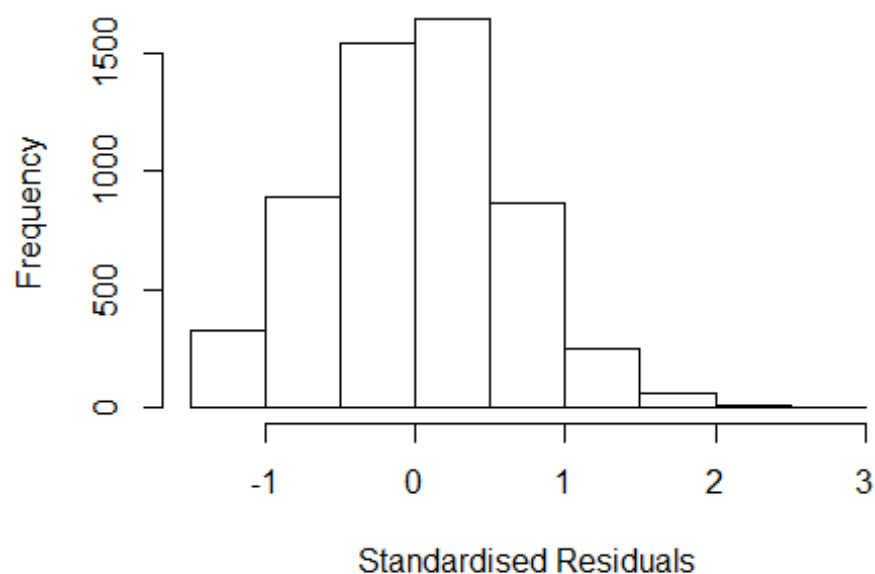
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1662 0033099611 3.971 1999      1706      3      3.088
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3859 -0.4164  0.0133  0.4147  3.0884
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.84882    0.05824   14.57  <2e-16 ***
## FirstAuthorFemale1 -0.02331    0.02188   -1.07  0.2867
## LastAuthorFemale1 -0.00632    0.02384   -0.27  0.7909
## UniqueAuthors2     0.21709    0.02270    9.56  <2e-16 ***
## UniqueAuthors3     0.25004    0.02464   10.15  <2e-16 ***
## UniqueAuthors4     0.33141    0.03053   10.85  <2e-16 ***
## UniqueAuthors5     0.33836    0.03107   10.89  <2e-16 ***
## Year1997          0.32214    0.16467    1.96  0.0505 .
## Year1998          0.08764    0.07087    1.24  0.2163
## Year1999          0.03380    0.07539    0.45  0.6539
```

```

## Year2000      -0.03843    0.07244   -0.53    0.5958
## Year2001      -0.05321    0.07160   -0.74    0.4574
## Year2002      -0.06787    0.07084   -0.96    0.3380
## Year2003       0.12173    0.06822    1.78    0.0744 .
## Year2004       0.15315    0.06885    2.22    0.0262 *
## Year2005       0.17516    0.06842    2.56    0.0105 *
## Year2006       0.20168    0.06651    3.03    0.0024 **
## Year2007       0.20564    0.06601    3.12    0.0018 **
## Year2008       0.13877    0.06502    2.13    0.0329 *
## Year2009       0.15943    0.06452    2.47    0.0135 *
## Year2010       0.12388    0.06418    1.93    0.0536 .
## Year2011       0.08926    0.06388    1.40    0.1624
## Year2012       0.05712    0.06333    0.90    0.3671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0626, Adjusted R-squared:  0.0589
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 508 is an outlier with |weight| = 0 ( < 1.8e-05);
## 448 weights are ~= 1. The remaining 5129 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.866  0.952  0.908  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.79e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
## factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.122 1 1.059
## LastAuthorFemale 1.122 1 1.059
## Year 1.024 16 1.001

```

## Residuals from first and last author



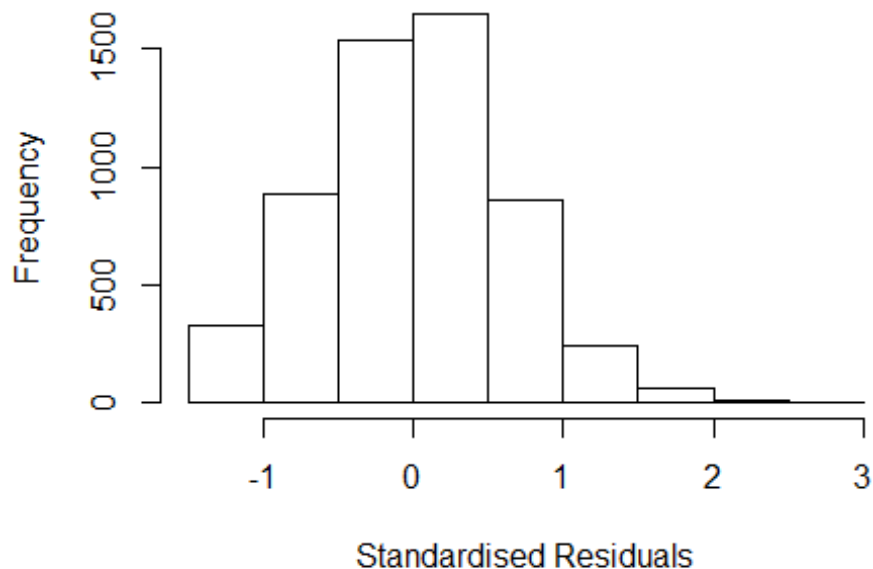
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1662 0033099611 3.971 1999      1706      3      2.968
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.27866 -0.42612  0.00986  0.43136  2.96770
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.956539   0.058590   16.33 < 2e-16 ***
## FirstAuthorFemale1 -0.021226   0.022330   -0.95  0.34186
## LastAuthorFemale1 -0.033139   0.024248   -1.37  0.17179
## Year1997         0.322121   0.168771    1.91  0.05636 .
## Year1998         0.119015   0.072222    1.65  0.09943 .
## Year1999         0.046757   0.077157    0.61  0.54454
## Year2000         0.006938   0.073812    0.09  0.92512
## Year2001        -0.000717   0.071916   -0.01  0.99204
## Year2002         0.005455   0.071829    0.08  0.93946
## Year2003         0.177831   0.069406    2.56  0.01043 *
## Year2004         0.213643   0.069914    3.06  0.00226 **
## Year2005         0.256931   0.069316    3.71  0.00021 ***
```

```

## Year2006          0.278772    0.067162    4.15  3.4e-05 ***
## Year2007          0.286659    0.065989    4.34  1.4e-05 ***
## Year2008          0.227480    0.065255    3.49  0.00049 ***
## Year2009          0.243589    0.064996    3.75  0.00018 ***
## Year2010          0.217422    0.064392    3.38  0.00074 ***
## Year2011          0.181989    0.064306    2.83  0.00467 **
## Year2012          0.145740    0.063987    2.28  0.02279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.0206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 508 is an outlier with |weight| = 0 ( < 1.8e-05);
## 466 weights are ~= 1. The remaining 5111 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0903 0.8660 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.79e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.011 1          1.005
## Year              1.011 16          1.000

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1662 0033099611 3.971 1999      1706      3      2.968
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.27484 -0.42719  0.00959  0.42974  2.97080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.951933   0.058402   16.30 < 2e-16 ***
## FirstAuthorFemale1 -0.034002   0.021396   -1.59  0.11208
## Year1997        0.322909   0.169083    1.91  0.05621 .
## Year1998        0.119330   0.072238    1.65  0.09861 .
## Year1999        0.048264   0.077146    0.63  0.53159
## Year2000        0.008693   0.073769    0.12  0.90619
## Year2001        0.000268   0.071917    0.00  0.99703
## Year2002        0.006905   0.071790    0.10  0.92338
## Year2003        0.179794   0.069411    2.59  0.00961 **
## Year2004        0.214523   0.069881    3.07  0.00215 **
## Year2005        0.258969   0.069251    3.74  0.00019 ***
## Year2006        0.280114   0.067123    4.17  3.1e-05 ***
```

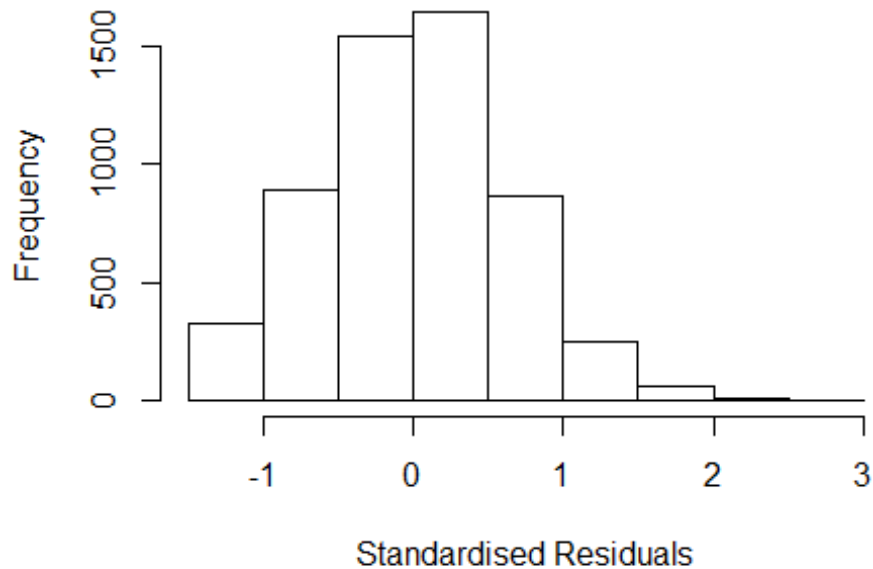


```

## Year2007          0.288254    0.065958    4.37  1.3e-05 ***
## Year2008          0.229555    0.065207    3.52  0.00043 ***
## Year2009          0.246228    0.064914    3.79  0.00015 ***
## Year2010          0.218734    0.064363    3.40  0.00068 ***
## Year2011          0.183107    0.064287    2.85  0.00441 **
## Year2012          0.147202    0.063951    2.30  0.02138 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.0234, Adjusted R-squared:  0.0205
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 508 is an outlier with |weight| = 0 ( < 1.8e-05);
## 466 weights are ~= 1. The remaining 5111 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0886 0.8650 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.79e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.011 1          1.005
## Year          1.011 16          1.000

```

## Residuals from last author



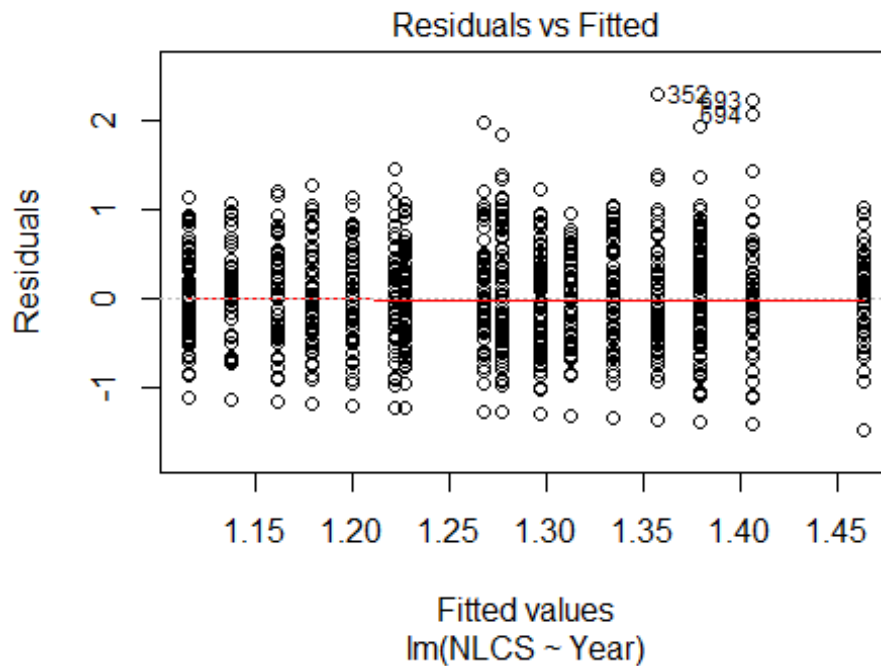
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1662 0033099611 3.971 1999      1706      3      2.968
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.274 -0.427  0.010  0.429  2.970
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95504    0.05859   16.30 < 2e-16 ***
## LastAuthorFemale1 -0.04243    0.02318   -1.83  0.06725 .
## Year1997        0.31900    0.16923    1.88  0.05949 .
## Year1998        0.11816    0.07224    1.64  0.10196
## Year1999        0.04605    0.07720    0.60  0.55086
## Year2000        0.00599    0.07384    0.08  0.93537
## Year2001       -0.00132    0.07197   -0.02  0.98533
## Year2002        0.00485    0.07188    0.07  0.94625
## Year2003        0.17664    0.06943    2.54  0.01098 *
## Year2004        0.21241    0.06996    3.04  0.00241 **
## Year2005        0.25541    0.06932    3.68  0.00023 ***
## Year2006        0.27799    0.06720    4.14  3.6e-05 ***
```

```

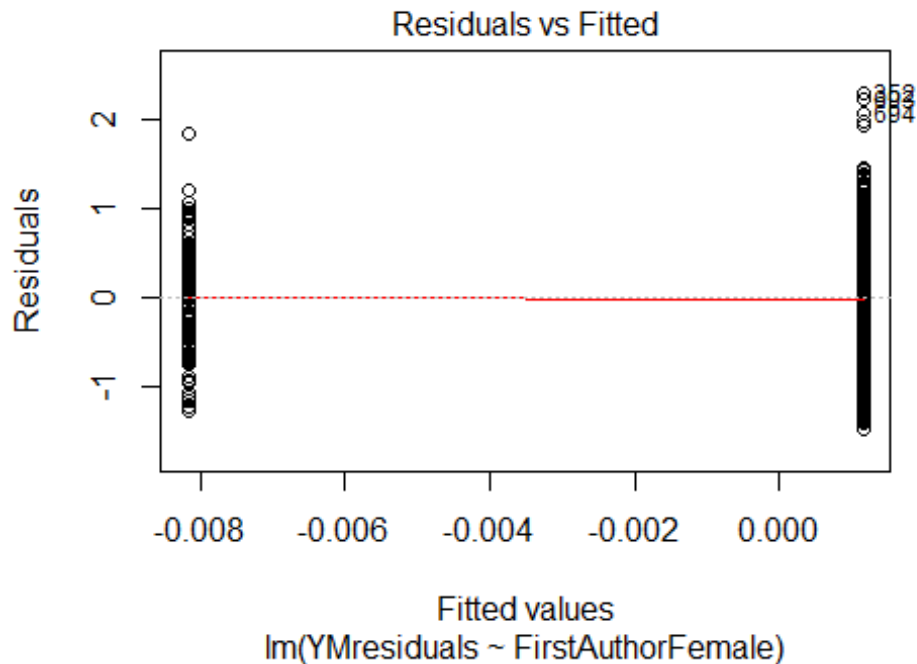
## Year2007          0.28593      0.06601      4.33  1.5e-05 ***
## Year2008          0.22572      0.06523      3.46  0.00054 ***
## Year2009          0.24117      0.06499      3.71  0.00021 ***
## Year2010          0.21611      0.06441      3.36  0.00080 ***
## Year2011          0.18078      0.06433      2.81  0.00497 **
## Year2012          0.14392      0.06398      2.25  0.02452 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.0206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 508 is an outlier with |weight| = 0 ( < 1.8e-05);
## 463 weights are ~= 1. The remaining 5114 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0899 0.8650 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.79e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5578"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1707"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   99  119  117  118  116  117  116  117  104  110  111  127  125  134  136
## 2011 2012
##  120  141
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   45   50   55   60   58   47   56   69   54   70   53   64   62   75   76

```

```
## 2011 2012
## 66 82
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 46 49 55 52 43 53 62 47 54 45 55 50 69 59
## 2011 2012
## 56 70
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 37, df = 16, p-value = 0.002
```

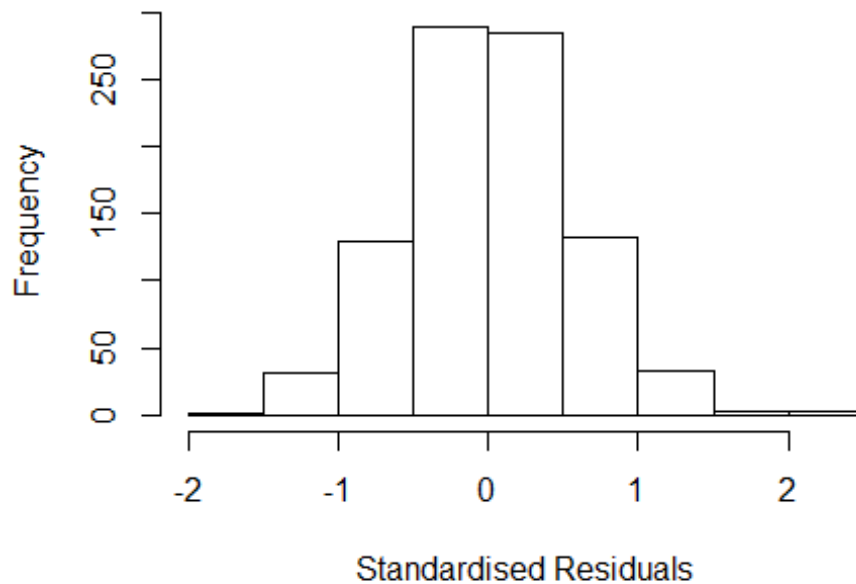


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.4520045653775"
## [1] "Male first author team size 2018 geometric mean: 2.84122338748731"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 310, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.64659128805981"
## [1] "Male last author team size 2018 geometric mean: 3.012981842937"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 270, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.190 1          1.091
## LastAuthorFemale  1.176 1          1.085
## UniqueAuthors    1.418 4          1.045
## Year             1.583 16          1.014
```

## Residuals from first and last author and team size



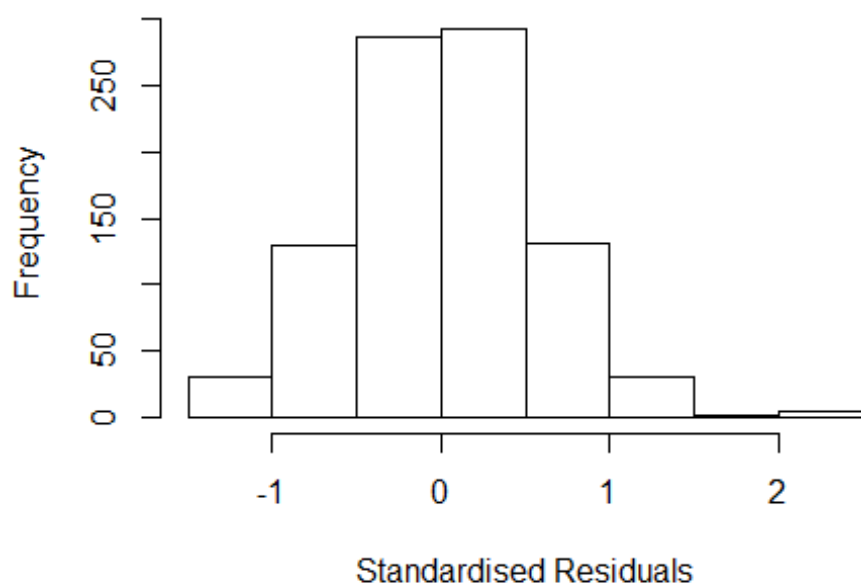
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.51367 -0.37916 0.00385 0.35733 2.36602
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3998 0.0906 15.45 < 2e-16 ***
## FirstAuthorFemale1 -0.0364 0.0592 -0.61 0.5389
## LastAuthorFemale1 0.0647 0.0572 1.13 0.2580
## UniqueAuthors2 0.1138 0.0555 2.05 0.0405 *
## UniqueAuthors3 0.1516 0.0590 2.57 0.0104 *
## UniqueAuthors4 0.2187 0.1044 2.10 0.0364 *
## UniqueAuthors5 0.3338 0.1034 3.23 0.0013 **
## Year1997 -0.1834 0.1314 -1.40 0.1633
## Year1998 -0.2227 0.1149 -1.94 0.0529 .
## Year1999 -0.2838 0.1150 -2.47 0.0138 *
```

```

## Year2000          -0.1631      0.1128    -1.45    0.1486
## Year2001          -0.2573      0.1339    -1.92    0.0549 .
## Year2002          -0.3135      0.1191    -2.63    0.0086 **
## Year2003          -0.2425      0.1127    -2.15    0.0317 *
## Year2004          -0.2903      0.1083    -2.68    0.0075 **
## Year2005          -0.3711      0.1161    -3.20    0.0014 **
## Year2006          -0.2142      0.1121    -1.91    0.0563 .
## Year2007          -0.3223      0.1124    -2.87    0.0042 **
## Year2008          -0.3343      0.1111    -3.01    0.0027 **
## Year2009          -0.4014      0.1019    -3.94    8.9e-05 ***
## Year2010          -0.3270      0.0993    -3.29    0.0010 **
## Year2011          -0.1353      0.1147    -1.18    0.2385
## Year2012          -0.3051      0.1265    -2.41    0.0161 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0442, Adjusted R-squared:  0.0204
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 831 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0266 0.8650 0.9520 0.9020 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.10e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.131 1 1.063
## LastAuthorFemale 1.182 1 1.087
## Year 1.190 16 1.005

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4912 -0.3913 0.0067 0.3646 2.3757
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4912 0.0808 18.45 < 2e-16 ***
## FirstAuthorFemale1 -0.0272 0.0576 -0.47 0.63672
## LastAuthorFemale1 0.0553 0.0567 0.97 0.32991
## Year1997 -0.1816 0.1307 -1.39 0.16496
## Year1998 -0.2098 0.1165 -1.80 0.07213 .
## Year1999 -0.2698 0.1149 -2.35 0.01907 *
## Year2000 -0.1428 0.1121 -1.27 0.20312
## Year2001 -0.2171 0.1325 -1.64 0.10171
## Year2002 -0.2785 0.1192 -2.34 0.01967 *
## Year2003 -0.2129 0.1116 -1.91 0.05688 .
## Year2004 -0.2732 0.1082 -2.53 0.01174 *
## Year2005 -0.3598 0.1150 -3.13 0.00182 **
```

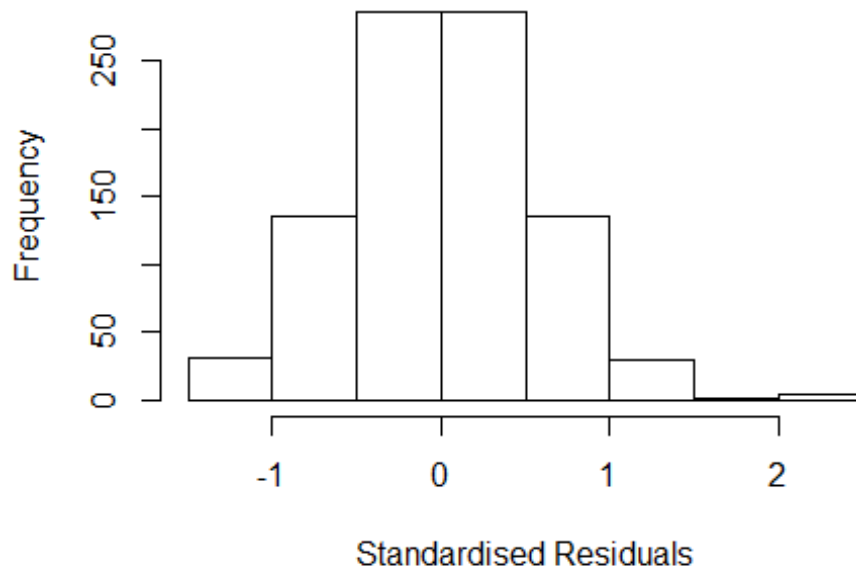


```

## Year2006          -0.1769      0.1114    -1.59   0.11268
## Year2007          -0.2830      0.1084    -2.61   0.00914 **
## Year2008          -0.3359      0.1118    -3.00   0.00274 **
## Year2009          -0.3689      0.1014    -3.64   0.00029 ***
## Year2010          -0.3016      0.0989    -3.05   0.00235 **
## Year2011          -0.0809      0.1146    -0.71   0.48045
## Year2012          -0.2580      0.1277    -2.02   0.04372 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0272, Adjusted R-squared:  0.0075
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 819 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0362 0.8660 0.9490 0.9030 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.074 1      1.036
## Year      1.074 16      1.002

```

## Residuals from first author



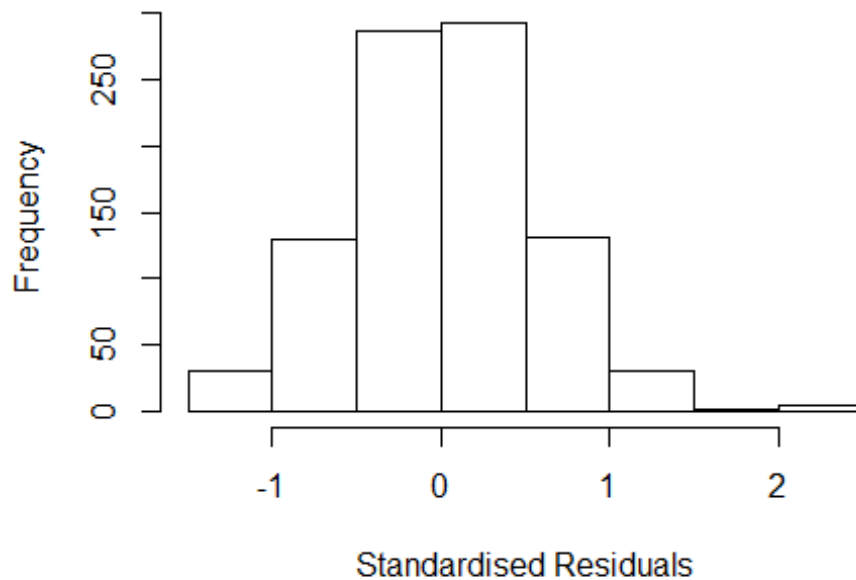
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.49176 -0.39677 0.00655 0.36675 2.37135
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4918 0.0810 18.41 < 2e-16 ***
## FirstAuthorFemale1 -0.0121 0.0566 -0.21 0.83011
## Year1997 -0.1795 0.1308 -1.37 0.17023
## Year1998 -0.2045 0.1162 -1.76 0.07890 .
## Year1999 -0.2693 0.1152 -2.34 0.01959 *
## Year2000 -0.1387 0.1122 -1.24 0.21702
## Year2001 -0.2161 0.1329 -1.63 0.10418
## Year2002 -0.2771 0.1190 -2.33 0.02012 *
## Year2003 -0.2100 0.1112 -1.89 0.05936 .
## Year2004 -0.2683 0.1080 -2.48 0.01320 *
## Year2005 -0.3526 0.1145 -3.08 0.00213 **
## Year2006 -0.1707 0.1113 -1.53 0.12537
```

```

## Year2007          -0.2822      0.1083   -2.61  0.00934 **
## Year2008          -0.3308      0.1115   -2.97  0.00308 **
## Year2009          -0.3647      0.1015   -3.60  0.00034 ***
## Year2010          -0.2998      0.0993   -3.02  0.00261 **
## Year2011          -0.0753      0.1139   -0.66  0.50875
## Year2012          -0.2473      0.1282   -1.93  0.05415 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.566
## Multiple R-squared:  0.0263, Adjusted R-squared:  0.00772
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 823 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0398 0.8650 0.9490 0.9040 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.126 1          1.061
## Year              1.126 16          1.004

```

## Residuals from last author



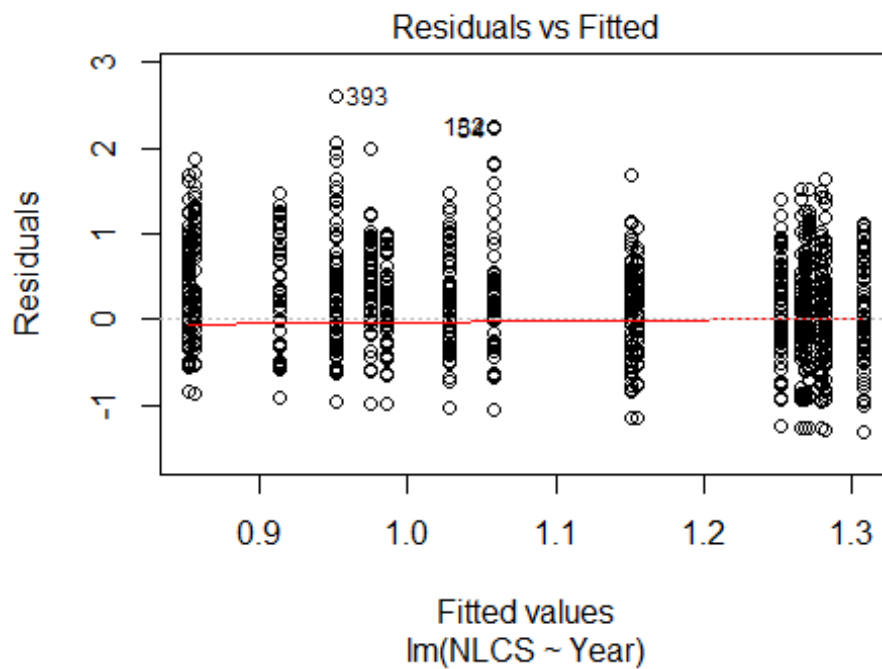
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.49009 -0.39559 0.00503 0.36741 2.37660
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4901 0.0810 18.39 < 2e-16 ***
## LastAuthorFemale1 0.0468 0.0554 0.85 0.39827
## Year1997 -0.1808 0.1308 -1.38 0.16712
## Year1998 -0.2087 0.1167 -1.79 0.07403 .
## Year1999 -0.2715 0.1151 -2.36 0.01851 *
## Year2000 -0.1427 0.1124 -1.27 0.20455
## Year2001 -0.2197 0.1323 -1.66 0.09725 .
## Year2002 -0.2800 0.1193 -2.35 0.01912 *
## Year2003 -0.2153 0.1114 -1.93 0.05348 .
## Year2004 -0.2751 0.1084 -2.54 0.01130 *
## Year2005 -0.3626 0.1148 -3.16 0.00165 **
## Year2006 -0.1802 0.1114 -1.62 0.10598
```

```

## Year2007          -0.2846      0.1084    -2.63   0.00881 **
## Year2008          -0.3383      0.1122    -3.01   0.00264 **
## Year2009          -0.3699      0.1016    -3.64   0.00029 ***
## Year2010          -0.3035      0.0992    -3.06   0.00229 **
## Year2011          -0.0838      0.1146    -0.73   0.46462
## Year2012          -0.2605      0.1277    -2.04   0.04175 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.563
## Multiple R-squared:  0.0269, Adjusted R-squared:  0.00838
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 823 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0351 0.8670 0.9500 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 909"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1708"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 127 144 169 175 238 182 121 100 91 109 136 137 134 115 140
## 2011 2012
## 112 113
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 67 97 91 129 80 64 67 56 71 76 76 77 72 85
## 2011 2012

```

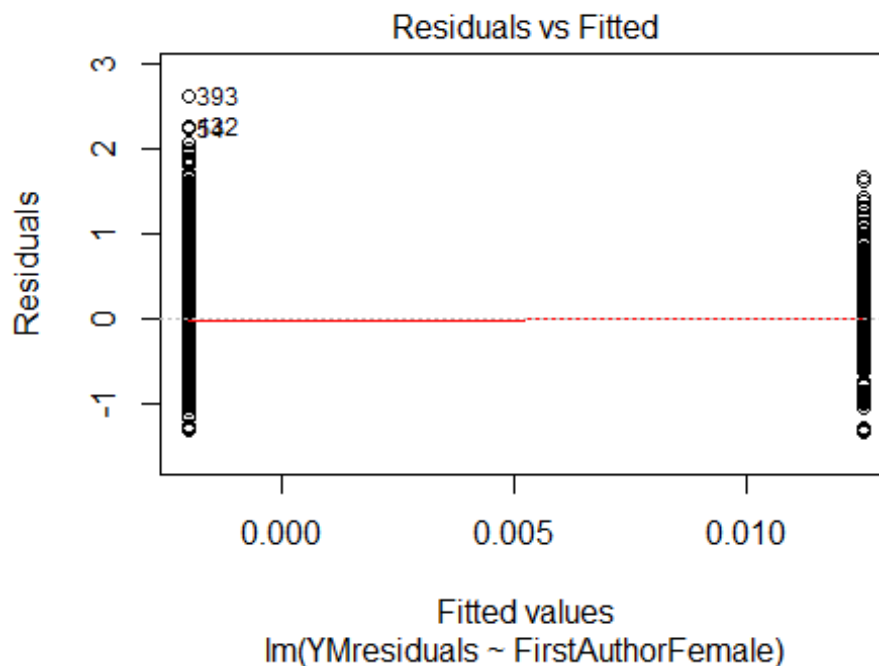
```
## 60 75
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 65 86 82 119 76 57 62 49 62 61 69 66 64 63
## 2011 2012
## 55 61
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 60, df = 16, p-value = 6e-07
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0038, df = 1, p-value = 1
## [1] "Female first author team size 2018 geometric mean: 3.44317533090674"
## [1] "Male first author team size 2018 geometric mean: 2.46714942723415"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

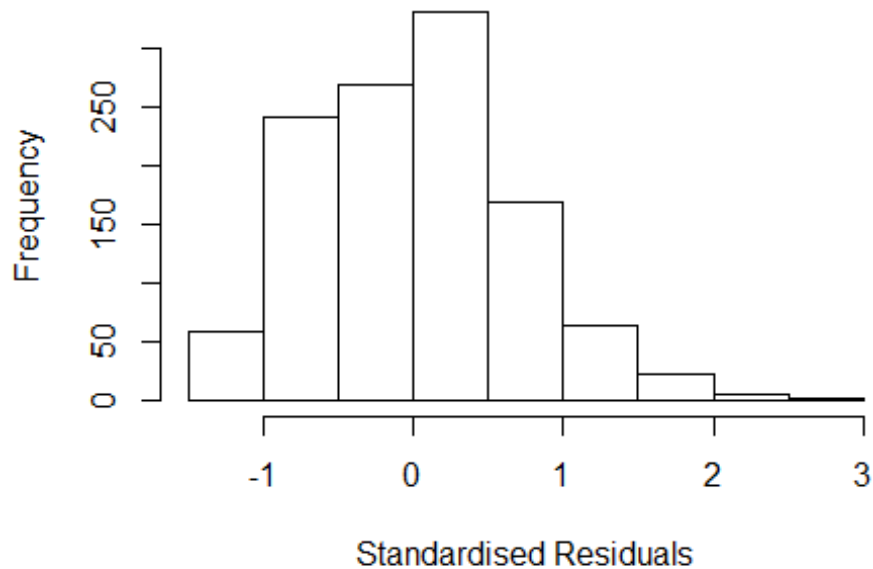
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 320, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.0861636884341"
## [1] "Male last author team size 2018 geometric mean: 2.58695280196159"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 160, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.326 1      1.151
## LastAuthorFemale 1.252 1      1.119
## UniqueAuthors    1.415 4      1.044
## Year              1.561 16     1.014
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 393 0031643297 3.557 1998    1705      4      2.922
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4631 -0.5244  0.0106  0.4261  2.9219
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7479    0.1139   6.57 7.8e-11 ***
## FirstAuthorFemale1  0.0169    0.0645   0.26  0.793
## LastAuthorFemale1 -0.0014    0.0732  -0.02  0.985
## UniqueAuthors2    0.4039    0.0564   7.16 1.5e-12 ***
## UniqueAuthors3    0.3845    0.0668   5.76 1.1e-08 ***
## UniqueAuthors4    0.5375    0.0771   6.97 5.5e-12 ***
## UniqueAuthors5    0.5029    0.0821   6.13 1.2e-09 ***
## Year1997         -0.1102    0.1438  -0.77  0.444
## Year1998         -0.1128    0.1399  -0.81  0.420
## Year1999         -0.1871    0.1516  -1.23  0.217
```

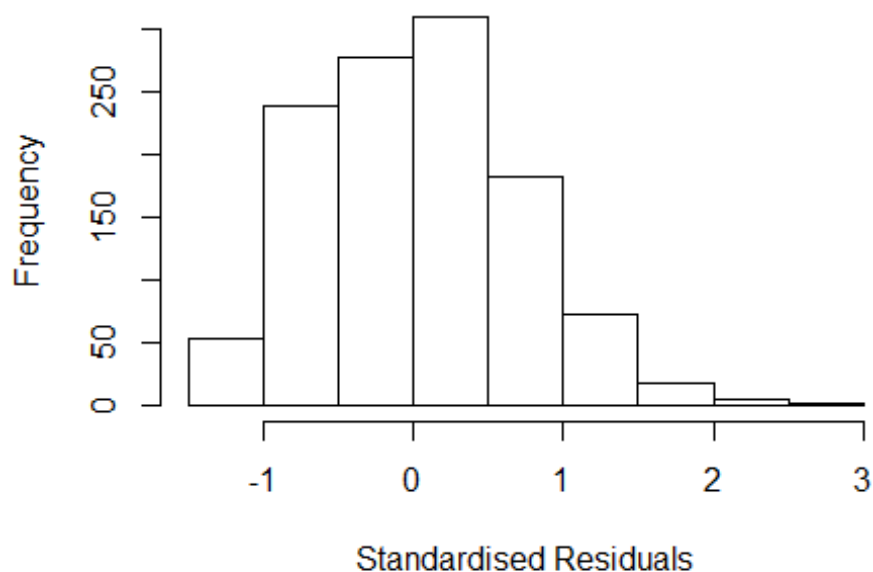


```

## Year2000          -0.1530      0.1267    -1.21     0.227
## Year2001          -0.1002      0.1457    -0.69     0.492
## Year2002           0.2450      0.1646     1.49     0.137
## Year2003           0.2544      0.1374     1.85     0.064 .
## Year2004           0.1606      0.1315     1.22     0.222
## Year2005           0.1406      0.1334     1.05     0.292
## Year2006           0.1453      0.1287     1.13     0.259
## Year2007           0.1777      0.1351     1.31     0.189
## Year2008           0.0162      0.1387     0.12     0.907
## Year2009           0.2188      0.1349     1.62     0.105
## Year2010          -0.0765      0.1427    -0.54     0.592
## Year2011          -0.1042      0.1355    -0.77     0.442
## Year2012           0.0438      0.1256     0.35     0.727
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.655
## Multiple R-squared:  0.14,   Adjusted R-squared:  0.123
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1049 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0089 0.8580 0.9370 0.9020 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.64e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.198 1          1.094
## LastAuthorFemale 1.121 1          1.059
## Year              1.126 16          1.004

```

## Residuals from first and last author



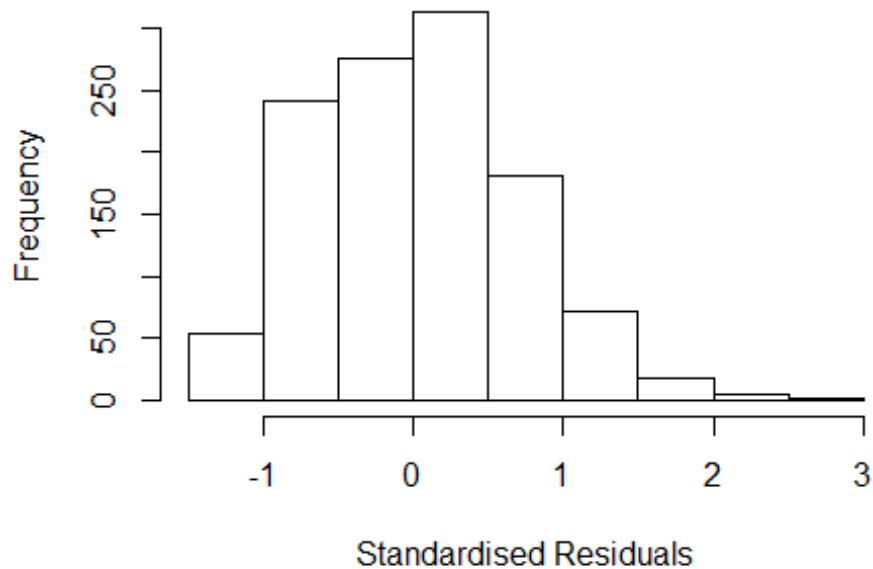
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 393 0031643297 3.557 1998      1705      4      2.719
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3275 -0.5033  0.0219  0.4708  2.7186
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9618    0.1165   8.26 4.2e-16 ***
## FirstAuthorFemale1  0.0403    0.0668   0.60  0.546
## LastAuthorFemale1 -0.0201    0.0774  -0.26  0.796
## Year1997        -0.0916    0.1544  -0.59  0.553
## Year1998        -0.1233    0.1488  -0.83  0.407
## Year1999        -0.1853    0.1614  -1.15  0.251
## Year2000        -0.1435    0.1344  -1.07  0.286
## Year2001        -0.0114    0.1540  -0.07  0.941
## Year2002         0.3140    0.1724   1.82  0.069 .
## Year2003         0.3455    0.1435   2.41  0.016 *
## Year2004         0.2824    0.1374   2.05  0.040 *
## Year2005         0.2649    0.1402   1.89  0.059 .
```

```

## Year2006          0.2657      0.1342      1.98      0.048 *
## Year2007          0.2868      0.1386      2.07      0.039 *
## Year2008          0.1581      0.1395      1.13      0.257
## Year2009          0.3328      0.1394      2.39      0.017 *
## Year2010          0.0630      0.1479      0.43      0.670
## Year2011          0.0269      0.1432      0.19      0.851
## Year2012          0.1871      0.1323      1.41      0.158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.701
## Multiple R-squared:  0.0655, Adjusted R-squared:  0.0507
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 1076 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0993 0.8740 0.9480 0.9100 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.088 1      1.043
## Year              1.088 16      1.003

```

## Residuals from first author



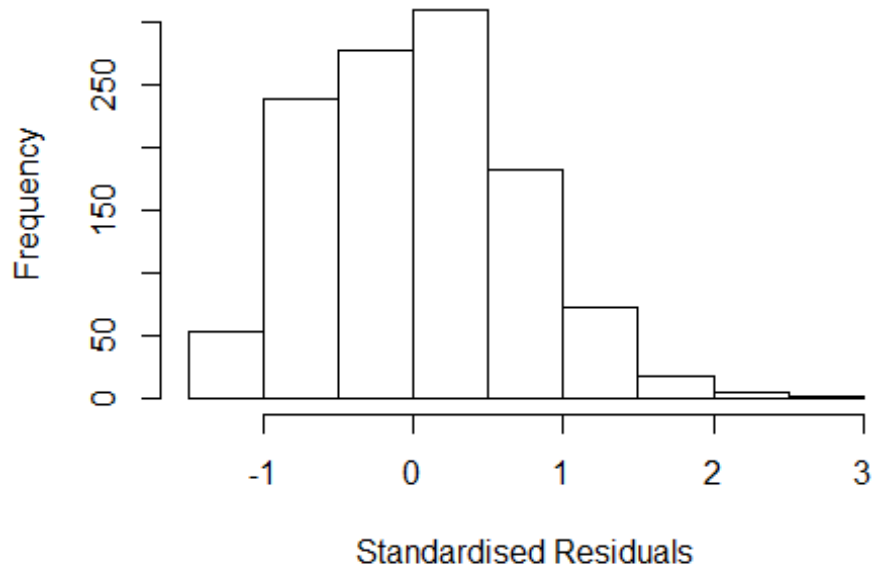
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 393 0031643297 3.557 1998    1705    4    2.719
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3408 -0.5064  0.0261  0.4693  2.7202
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9613    0.1165   8.25 4.4e-16 ***
## FirstAuthorFemale1 0.0337    0.0642   0.52  0.600
## Year1997      -0.0929    0.1541  -0.60  0.547
## Year1998      -0.1245    0.1488  -0.84  0.403
## Year1999      -0.1862    0.1613  -1.15  0.249
## Year2000      -0.1446    0.1327  -1.09  0.276
## Year2001      -0.0126    0.1537  -0.08  0.935
## Year2002       0.3125    0.1721   1.82  0.070 .
## Year2003       0.3459    0.1434   2.41  0.016 *
## Year2004       0.2814    0.1374   2.05  0.041 *
## Year2005       0.2624    0.1395   1.88  0.060 .
## Year2006       0.2661    0.1342   1.98  0.048 *
```

```

## Year2007          0.2865      0.1385      2.07      0.039 *
## Year2008          0.1580      0.1394      1.13      0.257
## Year2009          0.3315      0.1390      2.39      0.017 *
## Year2010          0.0612      0.1474      0.41      0.678
## Year2011          0.0253      0.1428      0.18      0.859
## Year2012          0.1869      0.1323      1.41      0.158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.0655, Adjusted R-squared:  0.0516
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 1077 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0969 0.8710 0.9480 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.02 1      1.010
## Year              1.02 16      1.001

```

## Residuals from last author



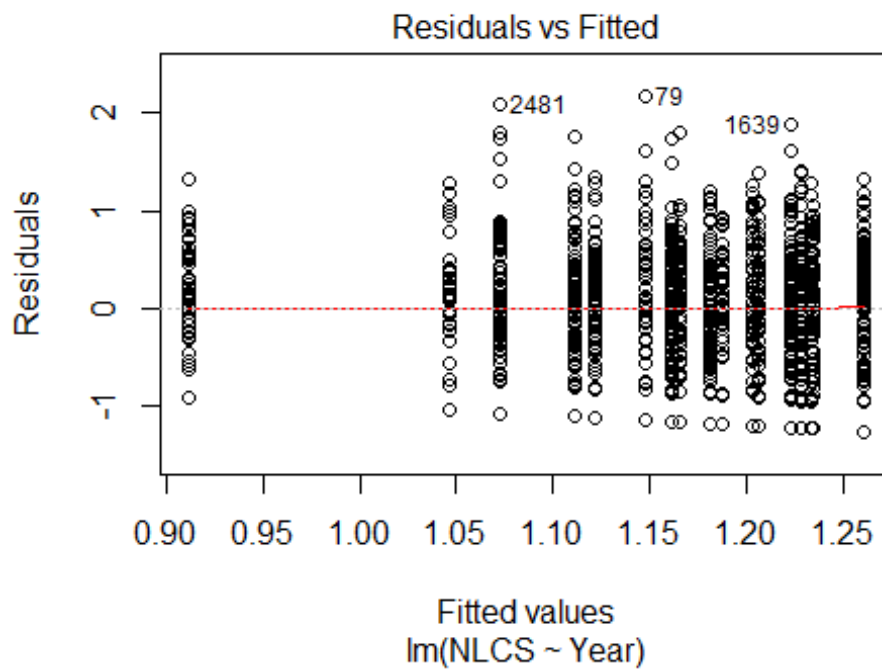
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 393 0031643297 3.557 1998    1705    4    2.719
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3116 -0.5076  0.0267  0.4648  2.7159
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96923    0.11481   8.44  <2e-16 ***
## LastAuthorFemale1 -0.00367    0.07352  -0.05   0.960
## Year1997      -0.09793    0.15353  -0.64   0.524
## Year1998      -0.12811    0.14854  -0.86   0.389
## Year1999      -0.18769    0.16169  -1.16   0.246
## Year2000      -0.14753    0.13390  -1.10   0.271
## Year2001      -0.01463    0.15344  -0.10   0.924
## Year2002       0.30814    0.17070   1.81   0.071 .
## Year2003       0.34233    0.14333   2.39   0.017 *
## Year2004       0.27868    0.13697   2.03   0.042 *
## Year2005       0.25712    0.13873   1.85   0.064 .
## Year2006       0.26595    0.13477   1.97   0.049 *
```

```

## Year2007      0.28362    0.13831    2.05    0.041 *
## Year2008      0.15669    0.13988    1.12    0.263
## Year2009      0.32616    0.13820    2.36    0.018 *
## Year2010      0.05686    0.14639    0.39    0.698
## Year2011      0.02714    0.14331    0.19    0.850
## Year2012      0.18631    0.13284    1.40    0.161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.702
## Multiple R-squared:  0.065, Adjusted R-squared:  0.0511
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 1079 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.101  0.873  0.948  0.910  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1158"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1709"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   75   73  101   89  118  125   76   71   94  116  159  197  157  154  137
## 2011 2012
##  162  170
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   48   75   52   59   81   45   49   68   80  119  142  109  111   90
## 2011 2012

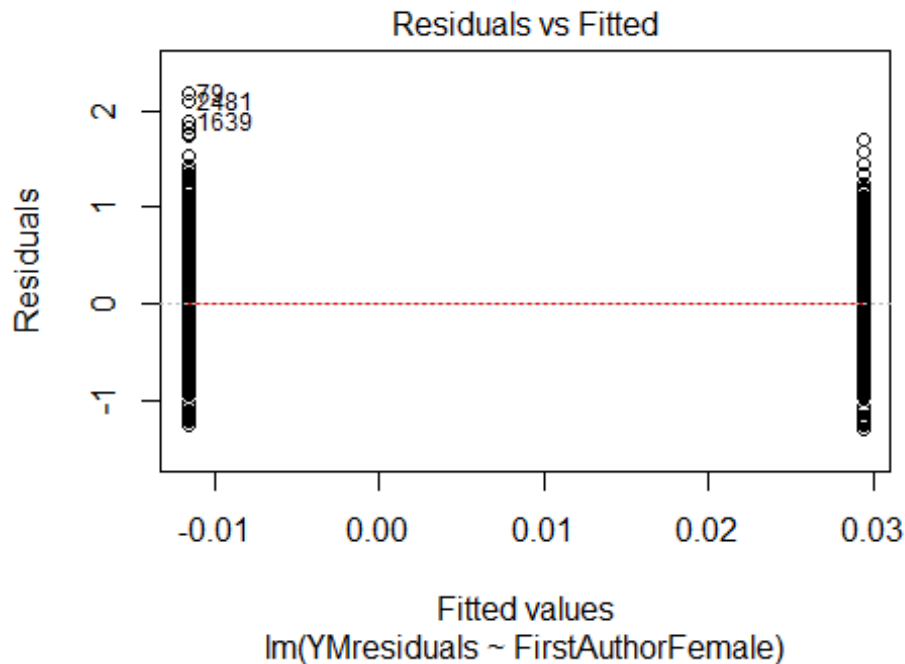
```

```
## 122 119
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 47 70 49 56 77 44 44 63 69 105 128 97 100 84
## 2011 2012
## 111 106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, df = 16, p-value = 0.2
```



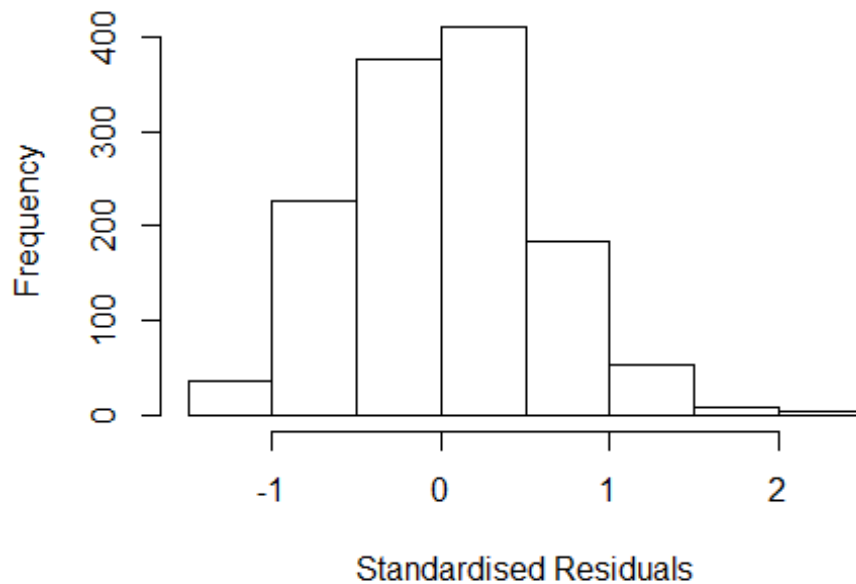
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.61, df = 1, p-value = 0.4
```





```
## [1] "Female first author team size 2018 geometric mean: 2.34247043076974"
## [1] "Male first author team size 2018 geometric mean: 2.28033499068246"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.08304783277192"
## [1] "Male last author team size 2018 geometric mean: 2.40696603725996"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.233 1      1.110
## LastAuthorFemale  1.208 1      1.099
## UniqueAuthors     1.333 4      1.037
## Year              1.437 16     1.011
```

## Residuals from first and last author and team size



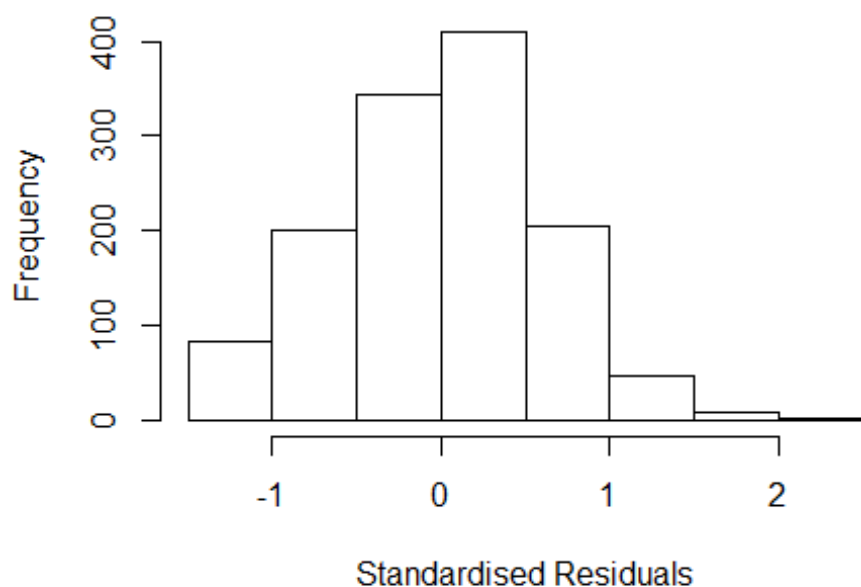
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3997 -0.3899 0.0158 0.3917 2.1648
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8287 0.1298 6.39 2.4e-10 ***
## FirstAuthorFemale1 0.0223 0.0415 0.54 0.59
## LastAuthorFemale1 0.0216 0.0420 0.52 0.61
## UniqueAuthors2 0.2998 0.0447 6.70 3.0e-11 ***
## UniqueAuthors3 0.4502 0.0475 9.48 < 2e-16 ***
## UniqueAuthors4 0.4070 0.0629 6.47 1.4e-10 ***
## UniqueAuthors5 0.4021 0.0638 6.30 4.1e-10 ***
## Year1997 -0.1286 0.1569 -0.82 0.41
## Year1998 0.1448 0.1488 0.97 0.33
## Year1999 0.0887 0.1582 0.56 0.57
```

```

## Year2000          0.1078      0.1536      0.70      0.48
## Year2001          0.0892      0.1459      0.61      0.54
## Year2002         -0.0607      0.1622     -0.37      0.71
## Year2003          0.0829      0.1492      0.56      0.58
## Year2004          0.0583      0.1516      0.38      0.70
## Year2005          0.1035      0.1447      0.72      0.47
## Year2006          0.1417      0.1417      1.00      0.32
## Year2007          0.0571      0.1375      0.42      0.68
## Year2008          0.1633      0.1398      1.17      0.24
## Year2009          0.0869      0.1417      0.61      0.54
## Year2010          0.0302      0.1399      0.22      0.83
## Year2011          0.0384      0.1380      0.28      0.78
## Year2012         -0.0430      0.1431     -0.30      0.76
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0885
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 110 weights are ~= 1. The remaining 1189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.130  0.868  0.949   0.906   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.157 1      1.076
## LastAuthorFemale  1.135 1      1.065
## Year              1.122 16      1.004

```

## Residuals from first and last author



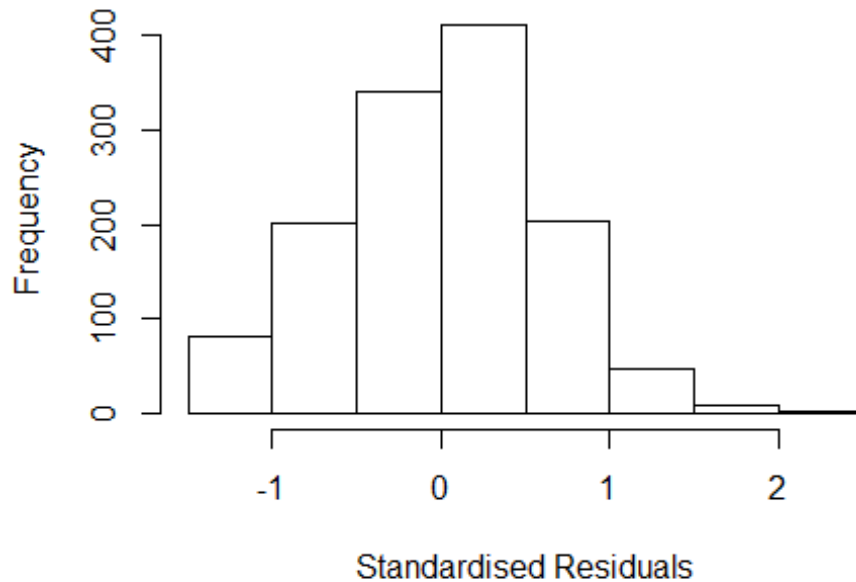
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2918 -0.4191 0.0279 0.4131 2.2684
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05522 0.13091 8.06 1.7e-15 ***
## FirstAuthorFemale1 0.04696 0.04282 1.10 0.27
## LastAuthorFemale1 -0.00859 0.04421 -0.19 0.85
## Year1997 -0.15050 0.16437 -0.92 0.36
## Year1998 0.14492 0.15215 0.95 0.34
## Year1999 0.13998 0.16119 0.87 0.39
## Year2000 0.18960 0.15913 1.19 0.23
## Year2001 0.10835 0.14813 0.73 0.46
## Year2002 -0.01052 0.16930 -0.06 0.95
## Year2003 0.12874 0.15514 0.83 0.41
## Year2004 0.10586 0.15469 0.68 0.49
## Year2005 0.10034 0.15148 0.66 0.51
```

```

## Year2006          0.17195      0.14716      1.17      0.24
## Year2007          0.10189      0.14038      0.73      0.47
## Year2008          0.13689      0.14298      0.96      0.34
## Year2009          0.11150      0.14526      0.77      0.44
## Year2010          0.03117      0.14324      0.22      0.83
## Year2011          0.06335      0.14193      0.45      0.66
## Year2012         -0.02767      0.14942     -0.19      0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.611
## Multiple R-squared:  0.0161, Adjusted R-squared:  0.00225
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1195 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.139  0.868  0.949  0.909  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1          1.041
## Year              1.084 16          1.003

```

## Residuals from first author



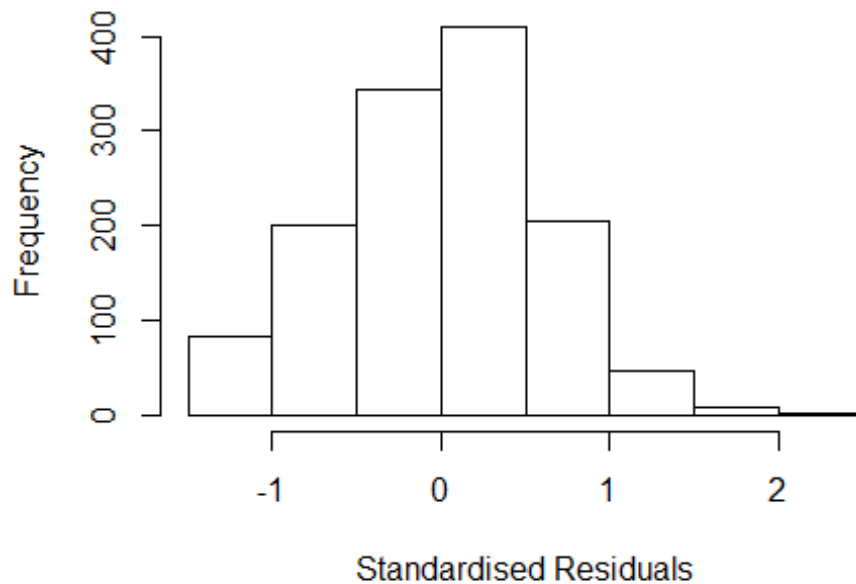
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2876 -0.4182 0.0285 0.4119 2.2607
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0543 0.1307 8.07 1.7e-15 ***
## FirstAuthorFemale1 0.0436 0.0419 1.04 0.30
## Year1997 -0.1502 0.1643 -0.91 0.36
## Year1998 0.1446 0.1520 0.95 0.34
## Year1999 0.1402 0.1611 0.87 0.38
## Year2000 0.1897 0.1590 1.19 0.23
## Year2001 0.1083 0.1480 0.73 0.46
## Year2002 -0.0106 0.1691 -0.06 0.95
## Year2003 0.1291 0.1550 0.83 0.40
## Year2004 0.1049 0.1544 0.68 0.50
## Year2005 0.1003 0.1514 0.66 0.51
## Year2006 0.1714 0.1470 1.17 0.24
```

```

## Year2007          0.1015      0.1402      0.72      0.47
## Year2008          0.1373      0.1429      0.96      0.34
## Year2009          0.1107      0.1449      0.76      0.45
## Year2010          0.0311      0.1431      0.22      0.83
## Year2011          0.0637      0.1418      0.45      0.65
## Year2012         -0.0278      0.1493     -0.19      0.85
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.612
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00299
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1190 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.143  0.868  0.949  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.063 1          1.031
## Year            1.063 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2513 -0.4210 0.0294 0.4080 2.2415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0618 0.1322 8.03 2.1e-15 ***
## LastAuthorFemale1 0.0117 0.0430 0.27 0.79
## Year1997 -0.1515 0.1653 -0.92 0.36
## Year1998 0.1445 0.1532 0.94 0.35
## Year1999 0.1419 0.1624 0.87 0.38
## Year2000 0.1859 0.1599 1.16 0.25
## Year2001 0.1080 0.1492 0.72 0.47
## Year2002 -0.0143 0.1698 -0.08 0.93
## Year2003 0.1283 0.1562 0.82 0.41
## Year2004 0.1055 0.1556 0.68 0.50
## Year2005 0.0980 0.1527 0.64 0.52
## Year2006 0.1778 0.1481 1.20 0.23
```

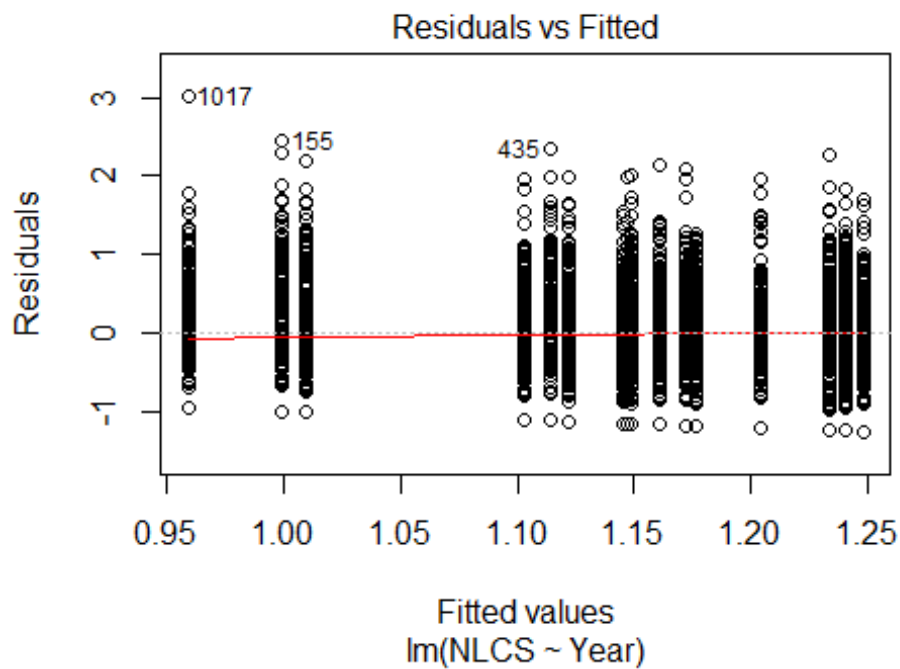


```

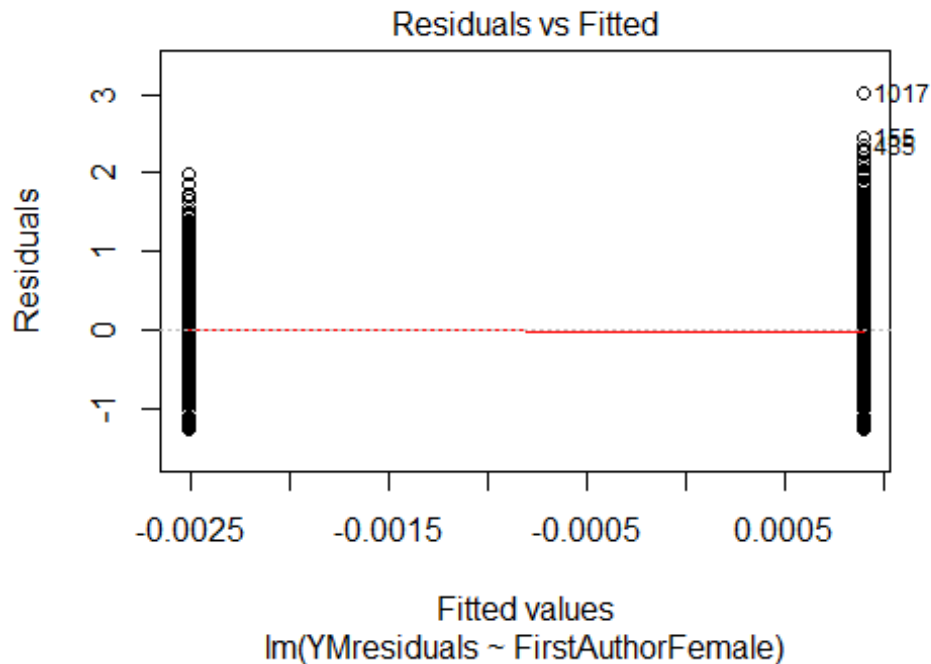
## Year2007          0.1057      0.1414      0.75      0.45
## Year2008          0.1405      0.1442      0.97      0.33
## Year2009          0.1151      0.1464      0.79      0.43
## Year2010          0.0348      0.1444      0.24      0.81
## Year2011          0.0669      0.1430      0.47      0.64
## Year2012         -0.0274      0.1505     -0.18      0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.00206
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1188 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.154  0.869  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1299"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1710"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  314  279  324  308  348  320  258  261  231  282  347  356  309  277  293
## 2011 2012
##  290  250
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  213  165  226  212  210  205  192  183  179  198  254  268  205  185  216
## 2011 2012

```

```
## 215 191
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 206 156 217 202 198 196 175 164 162 178 223 248 177 163 184
## 2011 2012
## 190 163
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 57, df = 16, p-value = 2e-06
```

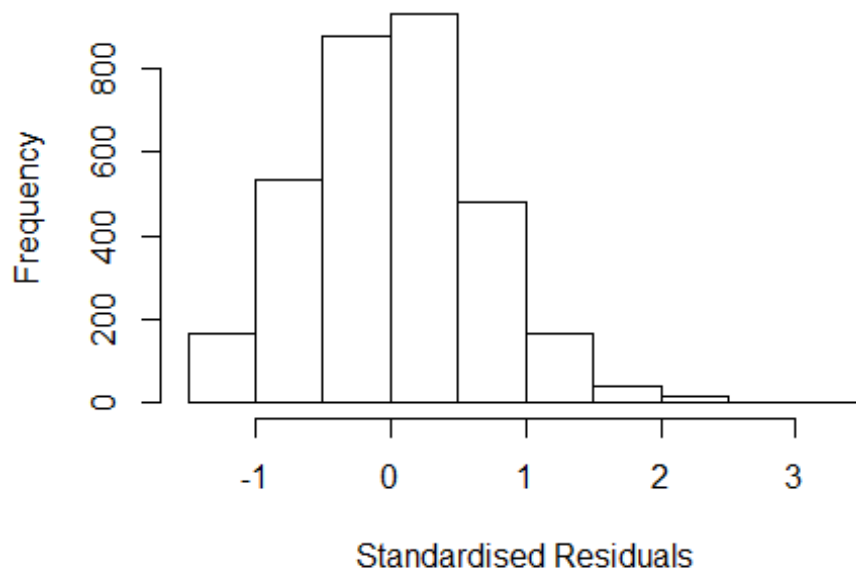


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.2, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.52579639448479"
## [1] "Male first author team size 2018 geometric mean: 2.09086978014213"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3600, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.00705141108472"
## [1] "Male last author team size 2018 geometric mean: 2.23467121062394"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.295 1          1.138
## LastAuthorFemale  1.284 1          1.133
## UniqueAuthors    1.207 4          1.024
## Year              1.253 16         1.007
```

## Residuals from first and last author and team size



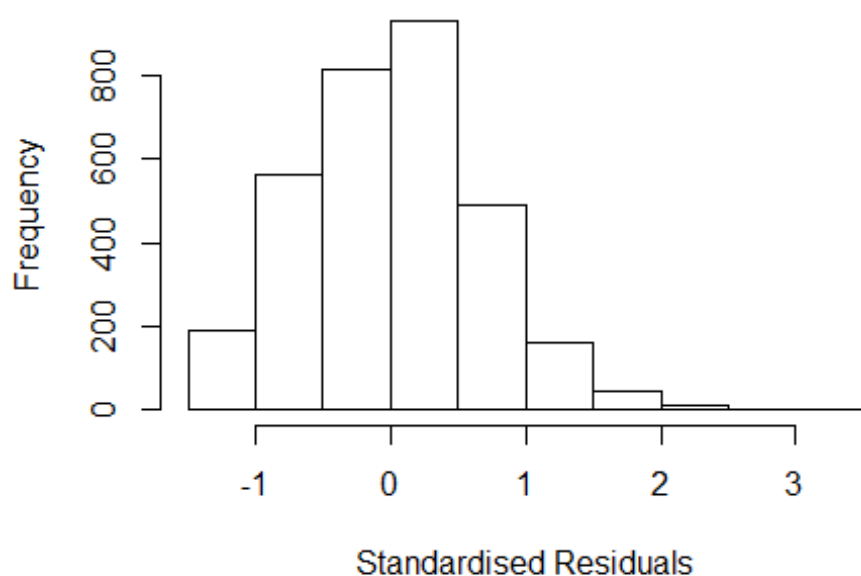
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1017 0033099611 3.971 1999      1706      3      3.174
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4574 -0.4314  0.0106  0.4258  3.1738
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.851585   0.058347  14.60 < 2e-16 ***
## FirstAuthorFemale1 0.000108   0.028601   0.00  0.99699
## LastAuthorFemale1 -0.003213   0.030649  -0.10  0.91651
## UniqueAuthors2    0.166729   0.027829   5.99  2.3e-09 ***
## UniqueAuthors3    0.306538   0.034683   8.84 < 2e-16 ***
## UniqueAuthors4    0.338586   0.050223   6.74  1.9e-11 ***
## UniqueAuthors5    0.294406   0.050090   5.88  4.6e-09 ***
## Year1997          0.096655   0.084268   1.15  0.25147
## Year1998          0.032413   0.075365   0.43  0.66716
## Year1999         -0.054349   0.076107  -0.71  0.47521
```

```

## Year2000      0.151504    0.073775    2.05    0.04010 *
## Year2001      0.131038    0.076936    1.70    0.08862 .
## Year2002      0.146986    0.079933    1.84    0.06603 .
## Year2003      0.188384    0.073858    2.55    0.01080 *
## Year2004      0.222853    0.077982    2.86    0.00429 **
## Year2005      0.148583    0.070745    2.10    0.03578 *
## Year2006      0.185345    0.071695    2.59    0.00978 **
## Year2007      0.251567    0.068819    3.66    0.00026 ***
## Year2008      0.270415    0.073453    3.68    0.00024 ***
## Year2009      0.155393    0.075716    2.05    0.04022 *
## Year2010      0.133281    0.070312    1.90    0.05811 .
## Year2011      0.197959    0.072674    2.72    0.00649 **
## Year2012      0.165158    0.074938    2.20    0.02760 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.0604, Adjusted R-squared:  0.0539
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 592 is an outlier with |weight| = 0 ( < 3.1e-05);
## 272 weights are ~1. The remaining 2929 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.118  0.868  0.951  0.908  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.12e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.276 1 1.129
## LastAuthorFemale 1.271 1 1.128
## Year 1.057 16 1.002

```

## Residuals from first and last author



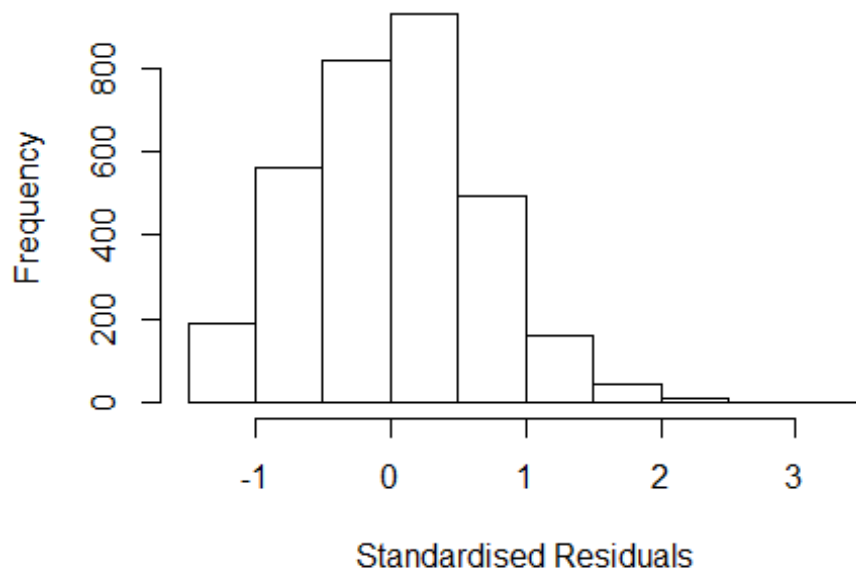
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 155  0000031451 3.437 1996    1600     4    2.501
## 1017 0033099611 3.971 1999    1706     3    3.073
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2702 -0.4466  0.0152  0.4424  3.0734
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9360    0.0593   15.79 < 2e-16 ***
## FirstAuthorFemale1  0.0139    0.0290    0.48  0.63050
## LastAuthorFemale1 -0.0240    0.0311   -0.77  0.43958
## Year1997         0.1173    0.0860    1.36  0.17257
## Year1998         0.0470    0.0774    0.61  0.54356
## Year1999        -0.0384    0.0793   -0.48  0.62874
## Year2000         0.1778    0.0751    2.37  0.01801 *
## Year2001         0.1467    0.0777    1.89  0.05916 .
## Year2002         0.1790    0.0802    2.23  0.02560 *
## Year2003         0.2520    0.0769    3.28  0.00107 **
## Year2004         0.2704    0.0805    3.36  0.00079 ***
```

```

## Year2005          0.1862      0.0725      2.57  0.01020 *
## Year2006          0.2397      0.0737      3.25  0.00117 **
## Year2007          0.3199      0.0696      4.60  4.5e-06 ***
## Year2008          0.3203      0.0739      4.34  1.5e-05 ***
## Year2009          0.2185      0.0766      2.85  0.00438 **
## Year2010          0.2020      0.0714      2.83  0.00470 **
## Year2011          0.2871      0.0729      3.94  8.3e-05 ***
## Year2012          0.2724      0.0748      3.64  0.00027 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0259, Adjusted R-squared:  0.0204
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 592 is an outlier with |weight| = 0 ( < 3.1e-05);
## 278 weights are ~= 1. The remaining 2923 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.105  0.867  0.947   0.907   0.986   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.12e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1           1.016
## Year              1.032 16           1.001

```

## Residuals from first author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 155  0000031451 3.437 1996    1600     4    2.501
## 1017 0033099611 3.971 1999    1706     3    3.073
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2563 -0.4490  0.0151  0.4382  3.0767
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93295    0.05911   15.78 < 2e-16 ***
## FirstAuthorFemale1 0.00217    0.02620    0.08  0.93405
## Year1997        0.11773    0.08608    1.37  0.17153
## Year1998        0.04772    0.07745    0.62  0.53786
## Year1999       -0.03865    0.07948   -0.49  0.62680
## Year2000        0.17790    0.07521    2.37  0.01807 *
## Year2001        0.14852    0.07761    1.91  0.05575 .
## Year2002        0.17998    0.08018    2.24  0.02486 *
## Year2003        0.25375    0.07681    3.30  0.00097 ***
## Year2004        0.26981    0.08061    3.35  0.00083 ***
## Year2005        0.18609    0.07254    2.57  0.01036 *
```

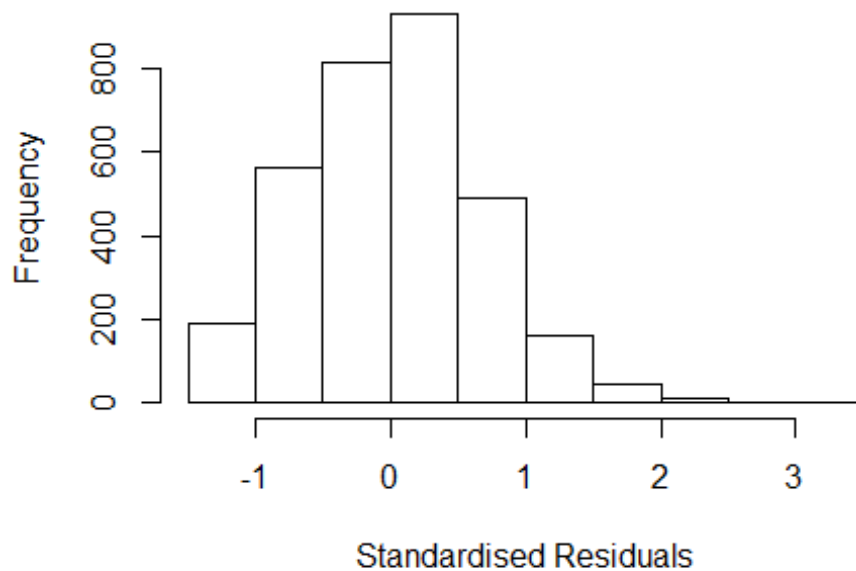


```

## Year2006          0.23987    0.07384    3.25  0.00117 **
## Year2007          0.31988    0.06977    4.58  4.7e-06 ***
## Year2008          0.32120    0.07392    4.35  1.4e-05 ***
## Year2009          0.21984    0.07663    2.87  0.00415 **
## Year2010          0.20262    0.07150    2.83  0.00463 **
## Year2011          0.28780    0.07289    3.95  8.0e-05 ***
## Year2012          0.27319    0.07489    3.65  0.00027 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.0206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 592 is an outlier with |weight| = 0 ( < 3.1e-05);
## 278 weights are ~ = 1. The remaining 2923 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.103  0.866   0.948   0.907   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.12e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.014
## Year            1.029 16          1.001

```

## Residuals from last author



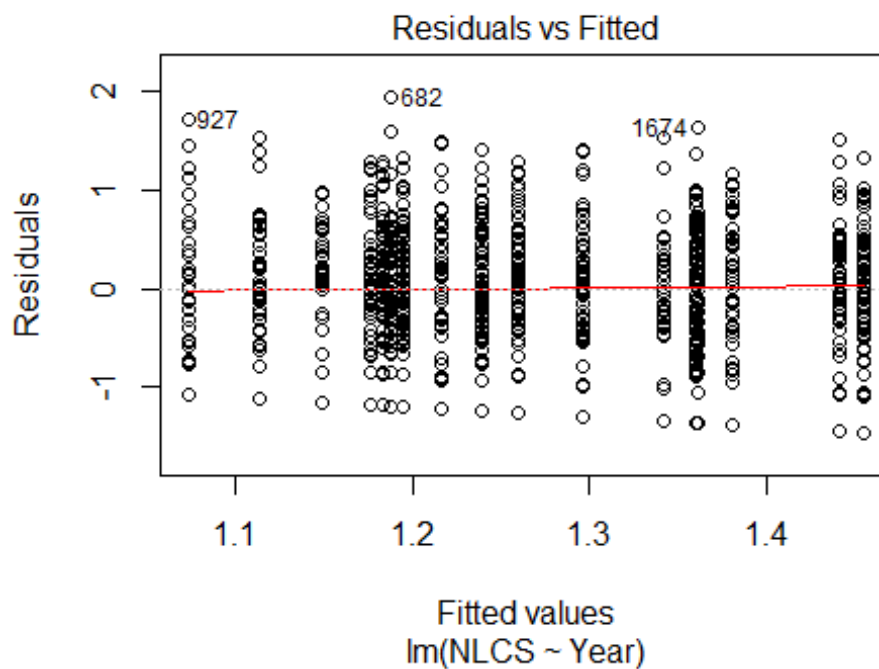
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 155  0000031451 3.437 1996    1600     4    2.501
## 1017 0033099611 3.971 1999    1706     3    3.073
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.259 -0.449  0.015  0.440  3.071
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9379    0.0592   15.86 < 2e-16 ***
## LastAuthorFemale1 -0.0166    0.0280   -0.59  0.55438
## Year1997        0.1165    0.0859    1.36  0.17527
## Year1998        0.0472    0.0774    0.61  0.54211
## Year1999       -0.0379    0.0793   -0.48  0.63272
## Year2000        0.1778    0.0751    2.37  0.01804 *
## Year2001        0.1472    0.0777    1.90  0.05815 .
## Year2002        0.1800    0.0801    2.25  0.02477 *
## Year2003        0.2519    0.0769    3.27  0.00107 **
## Year2004        0.2697    0.0805    3.35  0.00082 ***
## Year2005        0.1861    0.0725    2.57  0.01028 *
```

```

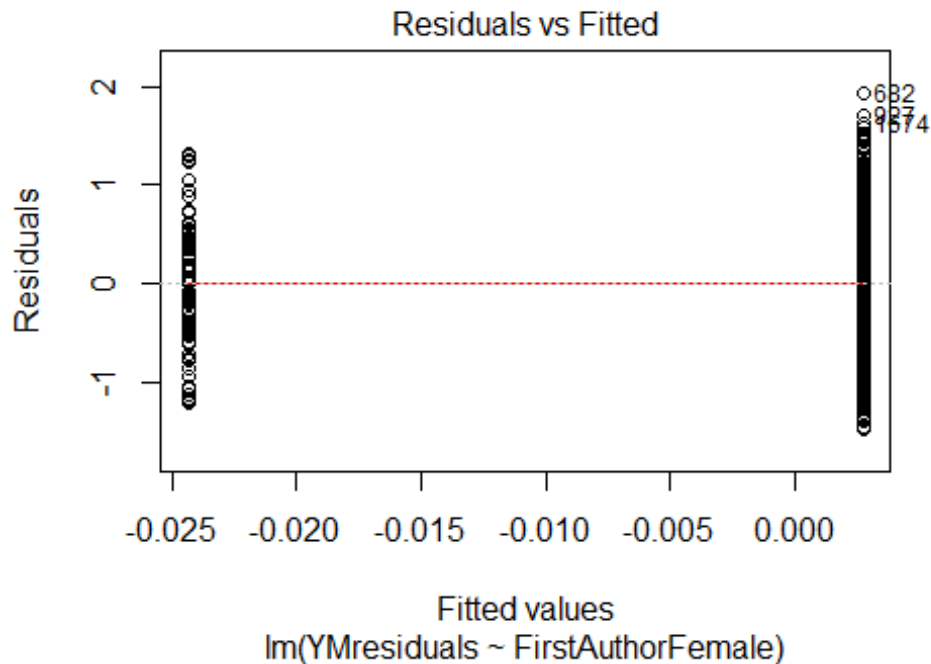
## Year2006          0.2400      0.0738      3.25  0.00115 **
## Year2007          0.3198      0.0697      4.59  4.6e-06 ***
## Year2008          0.3207      0.0739      4.34  1.5e-05 ***
## Year2009          0.2192      0.0766      2.86  0.00424 **
## Year2010          0.2021      0.0714      2.83  0.00469 **
## Year2011          0.2871      0.0729      3.94  8.4e-05 ***
## Year2012          0.2720      0.0748      3.64  0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.0259, Adjusted R-squared:  0.0207
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 592 is an outlier with |weight| = 0 ( < 3.1e-05);
## 281 weights are ~ = 1. The remaining 2920 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.105  0.866   0.948   0.907   0.986   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.12e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3202"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1711"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 127 136 156 141 141 128 116 110 94 109 157 189 152 141 176
## 2011 2012
## 163 174
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 50 41 50 57 56 39 46 45 44 57 51 82 58 64 88
## 2011 2012
## 81 76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 40 44 53 54 34 41 38 38 46 45 66 42 54 64
## 2011 2012
## 73 65
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 7.5, df = 16, p-value = 1
```

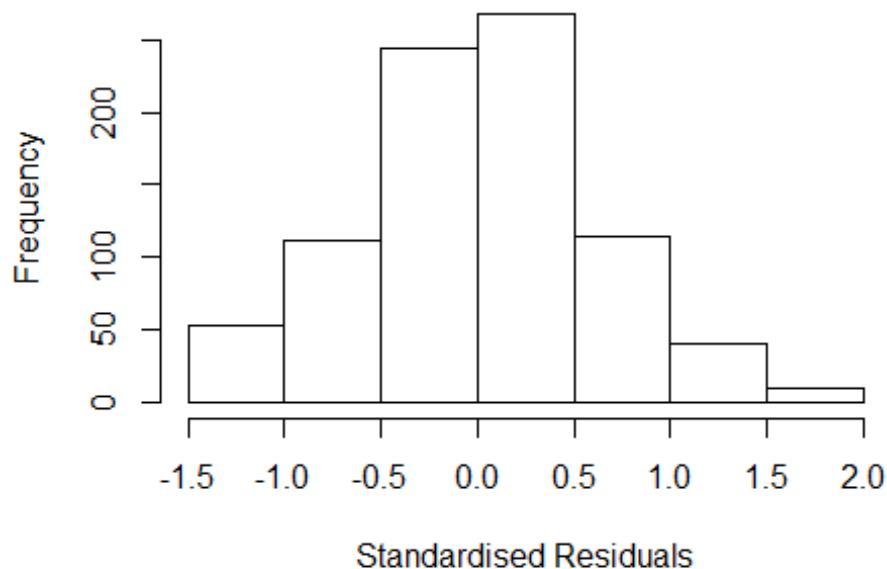


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.018, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 2.43071057062036"
## [1] "Male first author team size 2018 geometric mean: 2.80781753656461"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 260, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.1133064125246"
## [1] "Male last author team size 2018 geometric mean: 2.71982236582736"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.194 1      1.093
## LastAuthorFemale  1.190 1      1.091
## UniqueAuthors    1.973 4      1.089
## Year              2.100 16     1.023
```

## Residuals from first and last author and team size



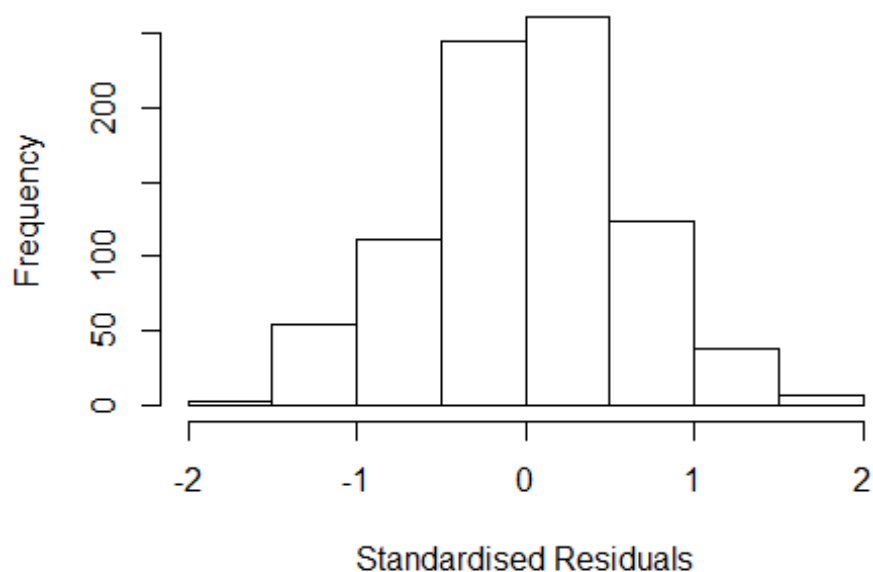
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.462 -0.400 0.016 0.399 1.908
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3432 0.1158 11.60 < 2e-16 ***
## FirstAuthorFemale1 -0.0257 0.0728 -0.35 0.72387
## LastAuthorFemale1 0.0398 0.0827 0.48 0.63089
## UniqueAuthors2 0.1657 0.0762 2.17 0.03007 *
## UniqueAuthors3 0.1739 0.0807 2.15 0.03152 *
## UniqueAuthors4 0.2392 0.1077 2.22 0.02659 *
## UniqueAuthors5 0.3946 0.1125 3.51 0.00047 ***
## Year1997 -0.1202 0.1333 -0.90 0.36774
## Year1998 -0.3786 0.1311 -2.89 0.00397 **
## Year1999 -0.3092 0.1261 -2.45 0.01443 *
```

```

## Year2000          -0.3071      0.1296    -2.37  0.01802 *
## Year2001          -0.4794      0.1603    -2.99  0.00287 **
## Year2002          -0.2836      0.1447    -1.96  0.05029 .
## Year2003          -0.1536      0.1491    -1.03  0.30308
## Year2004          -0.2629      0.1410    -1.86  0.06257 .
## Year2005          -0.3118      0.1338    -2.33  0.02005 *
## Year2006          -0.0927      0.1255    -0.74  0.46034
## Year2007          -0.2579      0.1197    -2.15  0.03152 *
## Year2008          -0.2465      0.1342    -1.84  0.06666 .
## Year2009          -0.1346      0.1342    -1.00  0.31619
## Year2010          -0.3818      0.1147    -3.33  0.00091 ***
## Year2011          -0.0223      0.1195    -0.19  0.85226
## Year2012          -0.2732      0.1246    -2.19  0.02863 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.0636, Adjusted R-squared:  0.0385
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 762 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.853   0.948   0.897   0.984   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           1.19e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.129 1           1.062
## LastAuthorFemale  1.147 1           1.071
## Year              1.183 16           1.005

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5120 -0.4051 0.0139 0.4150 1.9668
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4900 0.0890 16.73 <2e-16 ***
## FirstAuthorFemale1 -0.0314 0.0729 -0.43 0.6662
## LastAuthorFemale1 0.0141 0.0828 0.17 0.8645
## Year1997 -0.1500 0.1260 -1.19 0.2340
## Year1998 -0.4030 0.1298 -3.10 0.0020 **
## Year1999 -0.3205 0.1255 -2.55 0.0108 *
## Year2000 -0.3388 0.1238 -2.74 0.0063 **
## Year2001 -0.4677 0.1600 -2.92 0.0036 **
## Year2002 -0.2775 0.1424 -1.95 0.0516 .
## Year2003 -0.1347 0.1520 -0.89 0.3756
## Year2004 -0.2585 0.1381 -1.87 0.0617 .
## Year2005 -0.2944 0.1354 -2.17 0.0299 *
```

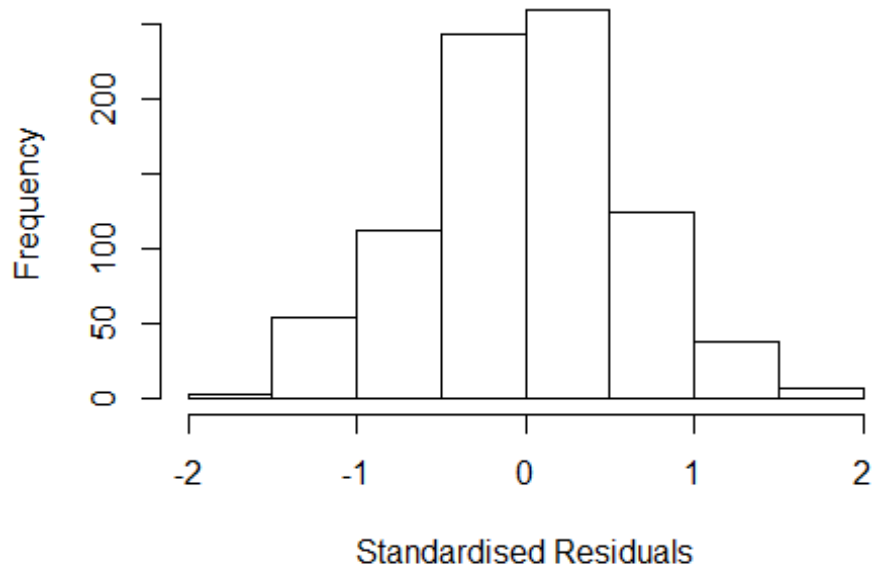


```

## Year2006          -0.0665      0.1264   -0.53   0.5990
## Year2007          -0.2403      0.1167   -2.06   0.0397 *
## Year2008          -0.2492      0.1302   -1.91   0.0560 .
## Year2009          -0.1118      0.1325   -0.84   0.3989
## Year2010          -0.3631      0.1117   -3.25   0.0012 **
## Year2011           0.0220      0.1159    0.19   0.8492
## Year2012          -0.2203      0.1228   -1.79   0.0732 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.605
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0246
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 767 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.268  0.867  0.948  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.099 1      1.048
## Year              1.099 16      1.003

```

## Residuals from first author



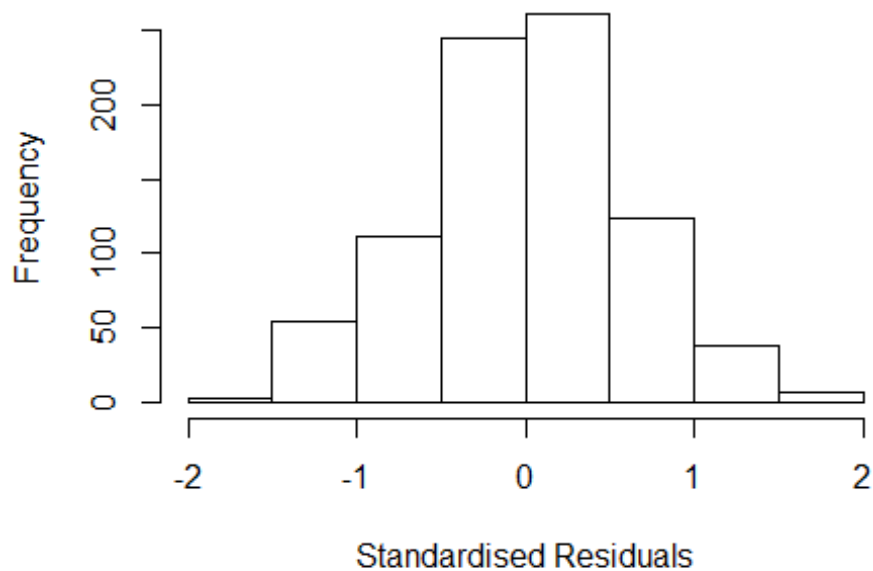
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5134 -0.4062 0.0131 0.4162 1.9664
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4898 0.0891 16.72 <2e-16 ***
## FirstAuthorFemale1 -0.0280 0.0720 -0.39 0.6976
## Year1997 -0.1492 0.1261 -1.18 0.2368
## Year1998 -0.4015 0.1295 -3.10 0.0020 **
## Year1999 -0.3200 0.1256 -2.55 0.0110 *
## Year2000 -0.3382 0.1238 -2.73 0.0064 **
## Year2001 -0.4674 0.1599 -2.92 0.0036 **
## Year2002 -0.2767 0.1424 -1.94 0.0524 .
## Year2003 -0.1343 0.1520 -0.88 0.3772
## Year2004 -0.2581 0.1382 -1.87 0.0622 .
## Year2005 -0.2934 0.1354 -2.17 0.0305 *
## Year2006 -0.0648 0.1259 -0.51 0.6068
```

```

## Year2007          -0.2388      0.1164   -2.05   0.0404 *
## Year2008          -0.2476      0.1299   -1.91   0.0571 .
## Year2009          -0.1112      0.1326   -0.84   0.4021
## Year2010          -0.3624      0.1117   -3.24   0.0012 **
## Year2011           0.0236      0.1154    0.20   0.8377
## Year2012          -0.2184      0.1219   -1.79   0.0736 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0456, Adjusted R-squared:  0.0259
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 766 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.267  0.866  0.948  0.902  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.116 1          1.057
## Year              1.116 16          1.003

```

## Residuals from last author



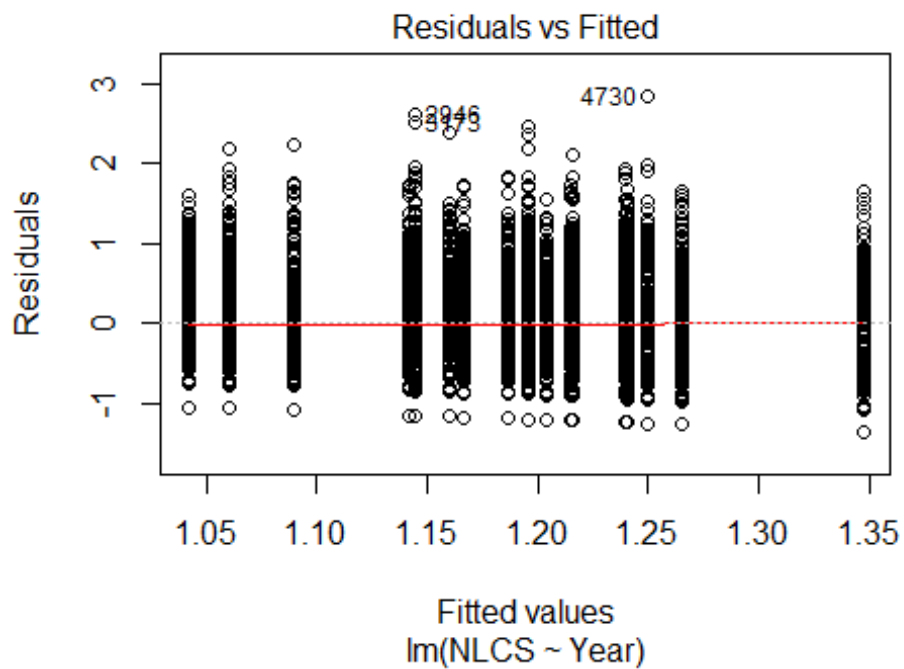
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5085 -0.4030 0.0146 0.4173 1.9680
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.48831 0.08965 16.60 <2e-16 ***
## LastAuthorFemale1 0.00444 0.08241 0.05 0.9570
## Year1997 -0.14958 0.12642 -1.18 0.2371
## Year1998 -0.40266 0.13013 -3.09 0.0020 **
## Year1999 -0.32018 0.12593 -2.54 0.0112 *
## Year2000 -0.33833 0.12421 -2.72 0.0066 **
## Year2001 -0.46709 0.16023 -2.92 0.0037 **
## Year2002 -0.27939 0.14239 -1.96 0.0501 .
## Year2003 -0.13578 0.15176 -0.89 0.3712
## Year2004 -0.25917 0.13817 -1.88 0.0611 .
## Year2005 -0.29557 0.13559 -2.18 0.0296 *
## Year2006 -0.07051 0.12598 -0.56 0.5759
```

```

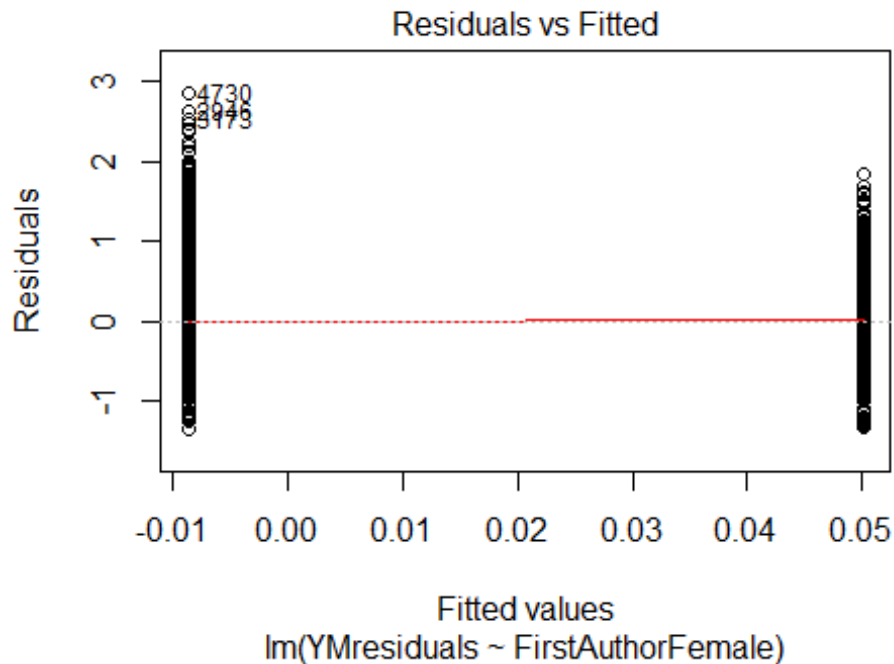
## Year2007          -0.24121      0.11685      -2.06      0.0393 *
## Year2008          -0.25091      0.13046      -1.92      0.0548 .
## Year2009          -0.11051      0.13305      -0.83      0.4064
## Year2010          -0.36557      0.11165      -3.27      0.0011 **
## Year2011           0.02018      0.11594       0.17      0.8619
## Year2012          -0.22228      0.12285      -1.81      0.0708 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0453, Adjusted R-squared:  0.0256
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 768 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.266  0.866  0.949   0.902   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 842"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1712"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  426  437  513  555  579  575  392  362  354  422  489  530  521  519  544
## 2011 2012
##  519  469
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  197  217  287  289  297  273  210  204  202  280  311  318  306  303  323
## 2011 2012

```

```
## 319 299
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 183 205 268 265 269 253 181 191 178 243 261 278 259 261 262
## 2011 2012
## 285 261
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 77, df = 16, p-value = 6e-10
```

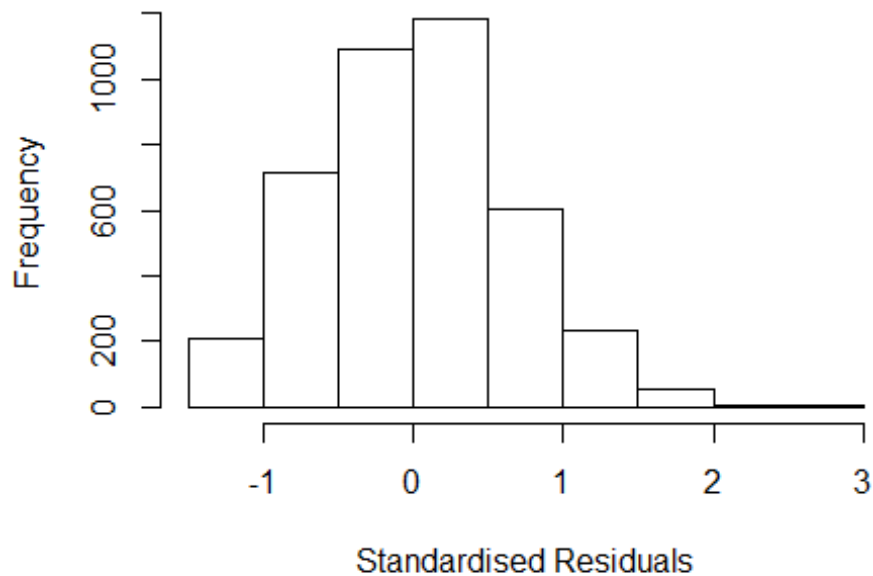


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.93906523297096"
## [1] "Male first author team size 2018 geometric mean: 2.59876394753585"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3900, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.87179595409468"
## [1] "Male last author team size 2018 geometric mean: 2.62600219826333"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.110 1      1.054
## LastAuthorFemale  1.123 1      1.060
## UniqueAuthors     1.125 4      1.015
## Year               1.144 16     1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1353  0031643297 3.557 1998    1705     4    2.579
## 2946  0001224048 3.749 2001    1702     4    2.860
## 4730  4344675264 4.089 2004    1704     3    2.781
## 8954  77749249761 3.553 2010    1712     3    2.666
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4855 -0.4336  0.0124  0.4337  2.8601
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.89607    0.06313   14.19  <2e-16 ***
## FirstAuthorFemale1 0.05233    0.03098    1.69   0.091 .
## LastAuthorFemale1 0.02292    0.03296    0.70   0.487
## UniqueAuthors2    0.26766    0.02691    9.95  <2e-16 ***
## UniqueAuthors3    0.31370    0.03083   10.18  <2e-16 ***
## UniqueAuthors4    0.40155    0.04051    9.91  <2e-16 ***
## UniqueAuthors5    0.42834    0.04656    9.20  <2e-16 ***
```

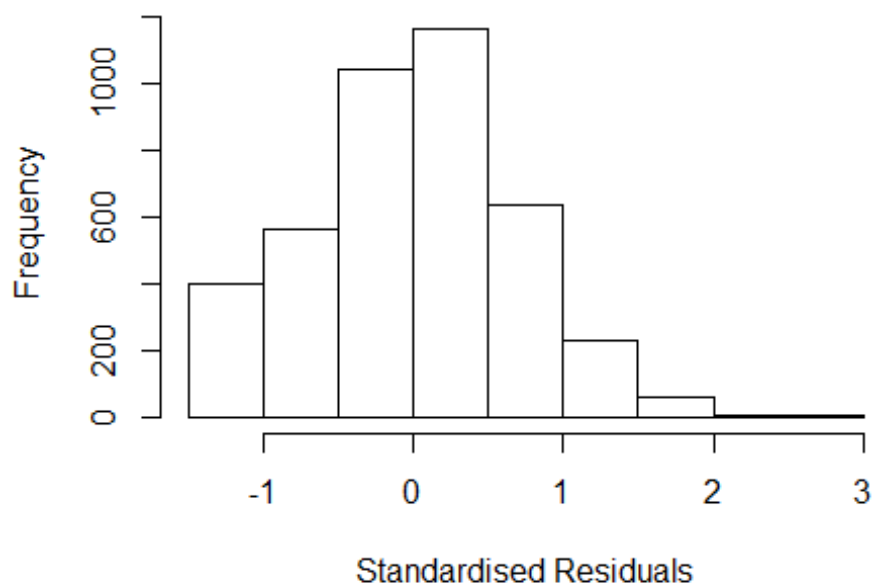


```

## Year1997          0.11895    0.08470    1.40    0.160
## Year1998          0.08174    0.07741    1.06    0.291
## Year1999         -0.04452    0.07894   -0.56    0.573
## Year2000         -0.04386    0.07703   -0.57    0.569
## Year2001         -0.00715    0.07947   -0.09    0.928
## Year2002          0.11442    0.08173    1.40    0.162
## Year2003          0.20518    0.07603    2.70    0.007 **
## Year2004          0.09782    0.08008    1.22    0.222
## Year2005          0.09037    0.07566    1.19    0.232
## Year2006          0.13553    0.07479    1.81    0.070 .
## Year2007          0.02937    0.07225    0.41    0.684
## Year2008          0.04296    0.07355    0.58    0.559
## Year2009          0.08664    0.07187    1.21    0.228
## Year2010         -0.00903    0.07263   -0.12    0.901
## Year2011         -0.02016    0.07292   -0.28    0.782
## Year2012          0.06972    0.07534    0.93    0.355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.659
## Multiple R-squared:  0.063, Adjusted R-squared:  0.058
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 372 weights are ~= 1. The remaining 3731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0199 0.8530 0.9500 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1      1.045
## LastAuthorFemale 1.103 1      1.050
## Year              1.023 16      1.001

```

## Residuals from first and last author



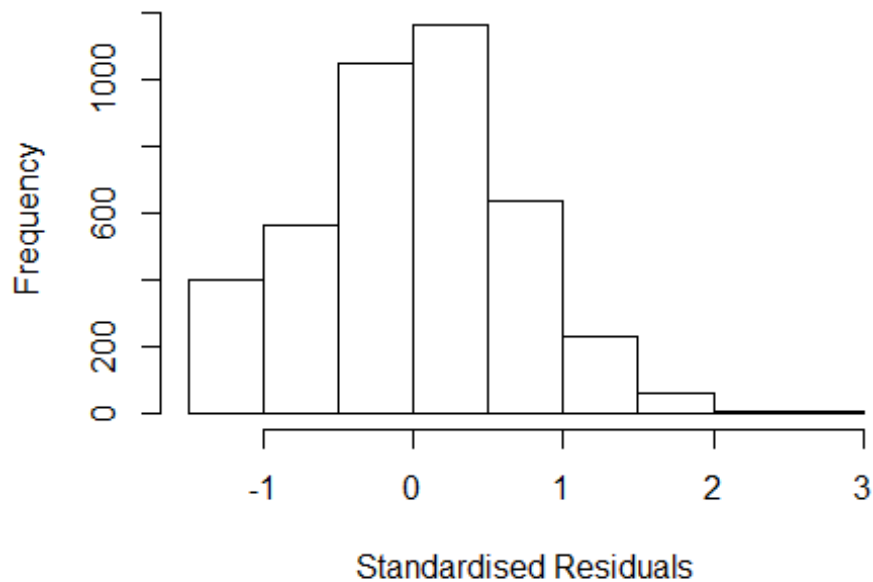
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1420 0032136153 3.657 1998    1702     3    2.503
## 2946 0001224048 3.749 2001    1702     4    2.673
## 3173 0035363218 3.647 2001    1702     5    2.571
## 4730 4344675264 4.089 2004    1704     3    2.887
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3262 -0.4610  0.0175  0.4516  2.8870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0434    0.0641   16.29 < 2e-16 ***
## FirstAuthorFemale1  0.0687    0.0312    2.20  0.02773 *
## LastAuthorFemale1  0.0100    0.0333    0.30  0.76310
## Year1997         0.1591    0.0852    1.87  0.06182 .
## Year1998         0.1101    0.0792    1.39  0.16480
## Year1999        -0.0156    0.0814   -0.19  0.84803
## Year2000        -0.0151    0.0798   -0.19  0.85002
## Year2001         0.0323    0.0815    0.40  0.69184
## Year2002         0.1724    0.0832    2.07  0.03829 *
```

```

## Year2003          0.2827      0.0778      3.64  0.00028 ***
## Year2004          0.1586      0.0817      1.94  0.05239 .
## Year2005          0.1533      0.0784      1.96  0.05039 .
## Year2006          0.2031      0.0764      2.66  0.00788 **
## Year2007          0.1017      0.0742      1.37  0.17080
## Year2008          0.1064      0.0749      1.42  0.15525
## Year2009          0.1554      0.0737      2.11  0.03509 *
## Year2010          0.0817      0.0741      1.10  0.26980
## Year2011          0.0781      0.0748      1.04  0.29630
## Year2012          0.1817      0.0765      2.38  0.01759 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.677
## Multiple R-squared:  0.0143, Adjusted R-squared:  0.00997
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 352 weights are ~= 1. The remaining 3751 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0295 0.8640 0.9490 0.9080 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.007 1          1.004
## Year              1.007 16          1.000

```

## Residuals from first author



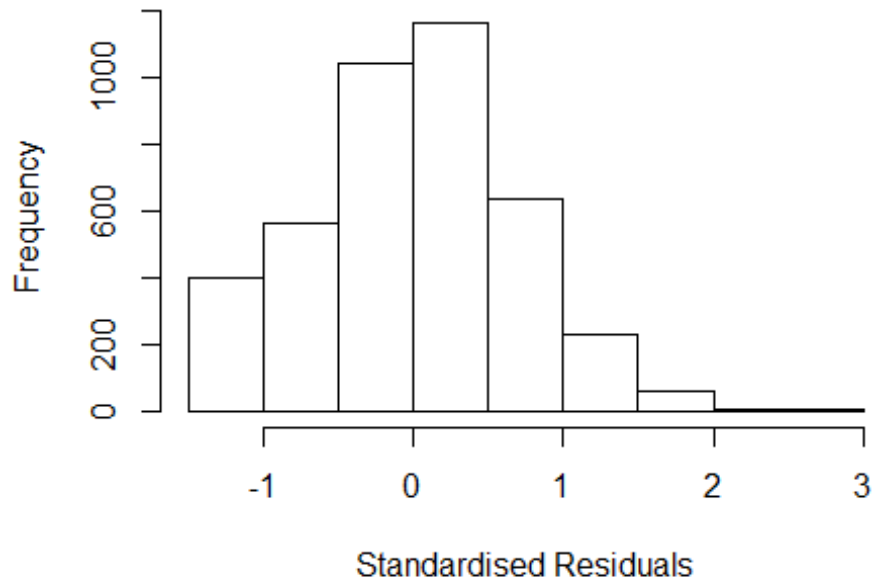
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1420 0032136153 3.657 1998    1702     3    2.503
## 2946 0001224048 3.749 2001    1702     4    2.673
## 3173 0035363218 3.647 2001    1702     5    2.571
## 4730 4344675264 4.089 2004    1704     3    2.887
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.327 -0.462  0.018  0.452  2.886
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0442    0.0640   16.30 < 2e-16 ***
## FirstAuthorFemale1  0.0721    0.0301    2.40  0.01658 *
## Year1997         0.1592    0.0852    1.87  0.06168 .
## Year1998         0.1102    0.0792    1.39  0.16415
## Year1999        -0.0154    0.0814   -0.19  0.84994
## Year2000        -0.0150    0.0798   -0.19  0.85084
## Year2001         0.0325    0.0815    0.40  0.68991
## Year2002         0.1728    0.0832    2.08  0.03784 *
## Year2003         0.2824    0.0778    3.63  0.00029 ***
```

```

## Year2004          0.1588      0.0817      1.94  0.05196 .
## Year2005          0.1540      0.0783      1.97  0.04927 *
## Year2006          0.2032      0.0764      2.66  0.00786 **
## Year2007          0.1017      0.0743      1.37  0.17073
## Year2008          0.1062      0.0749      1.42  0.15622
## Year2009          0.1555      0.0737      2.11  0.03500 *
## Year2010          0.0818      0.0741      1.10  0.26923
## Year2011          0.0782      0.0748      1.05  0.29544
## Year2012          0.1817      0.0765      2.38  0.01758 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.677
## Multiple R-squared:  0.0143, Adjusted R-squared:  0.0102
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 349 weights are ~= 1. The remaining 3754 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0299 0.8640 0.9490 0.9080 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.019 1      1.010
## Year      1.019 16      1.001

```

## Residuals from last author



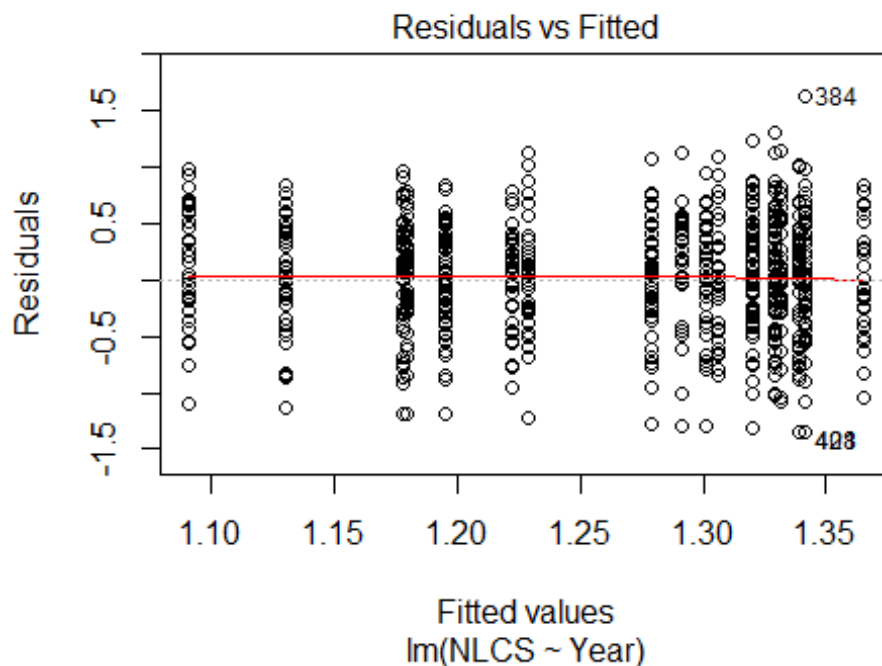
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1420 0032136153 3.657 1998    1702     3    2.503
## 2946 0001224048 3.749 2001    1702     4    2.673
## 3173 0035363218 3.647 2001    1702     5    2.571
## 4730 4344675264 4.089 2004    1704     3    2.887
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3332 -0.4639  0.0161  0.4558  2.8777
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0501     0.0640   16.42 < 2e-16 ***
## LastAuthorFemale1  0.0356     0.0320    1.11  0.26517
## Year1997          0.1592     0.0853    1.87  0.06206 .
## Year1998          0.1086     0.0794    1.37  0.17114
## Year1999         -0.0163     0.0814   -0.20  0.84169
## Year2000         -0.0160     0.0798   -0.20  0.84085
## Year2001          0.0332     0.0816    0.41  0.68367
## Year2002          0.1713     0.0831    2.06  0.03935 *
## Year2003          0.2830     0.0778    3.64  0.00028 ***
```

```

## Year2004          0.1611      0.0817      1.97  0.04860 *
## Year2005          0.1505      0.0783      1.92  0.05453 .
## Year2006          0.2067      0.0763      2.71  0.00681 **
## Year2007          0.1022      0.0742      1.38  0.16887
## Year2008          0.1066      0.0749      1.42  0.15459
## Year2009          0.1571      0.0737      2.13  0.03309 *
## Year2010          0.0828      0.0740      1.12  0.26334
## Year2011          0.0811      0.0747      1.09  0.27794
## Year2012          0.1812      0.0765      2.37  0.01790 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.678
## Multiple R-squared:  0.0131, Adjusted R-squared:  0.00904
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 347 weights are ~= 1. The remaining 3756 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0323 0.8630 0.9500 0.9080 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.44e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4103"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1800"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   51   39   47   63   59   68   66   46   34   51   54   79   58   73   59
## 2011 2012
##   55   75
##

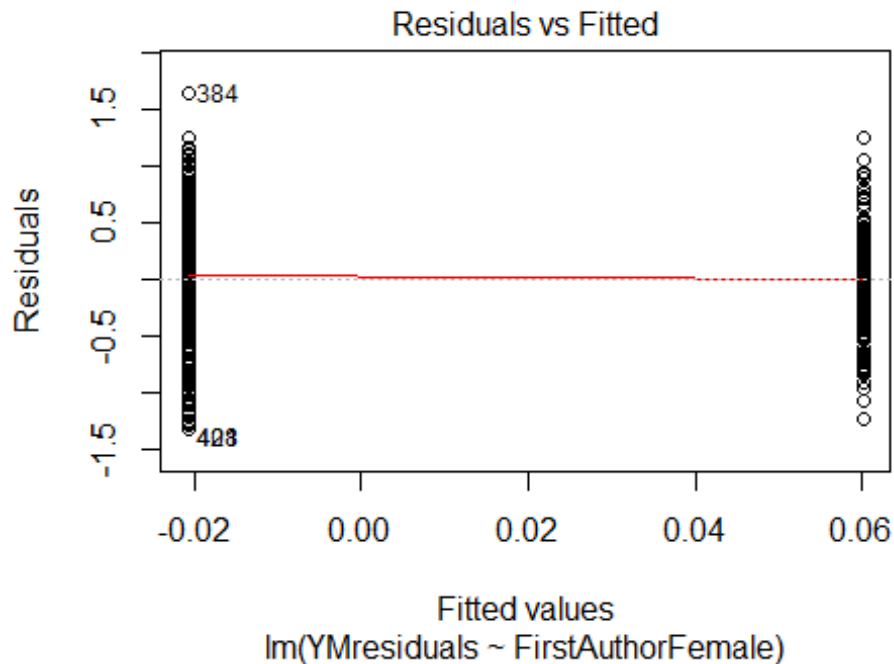
```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    42    25    28    40    44    43    49    37    27    44    42    66    50    61    49
## 2011 2012
##    44    57
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    40    24    26    39    40    41    47    31    26    41    40    59    48    57    39
## 2011 2012
##    41    54
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```



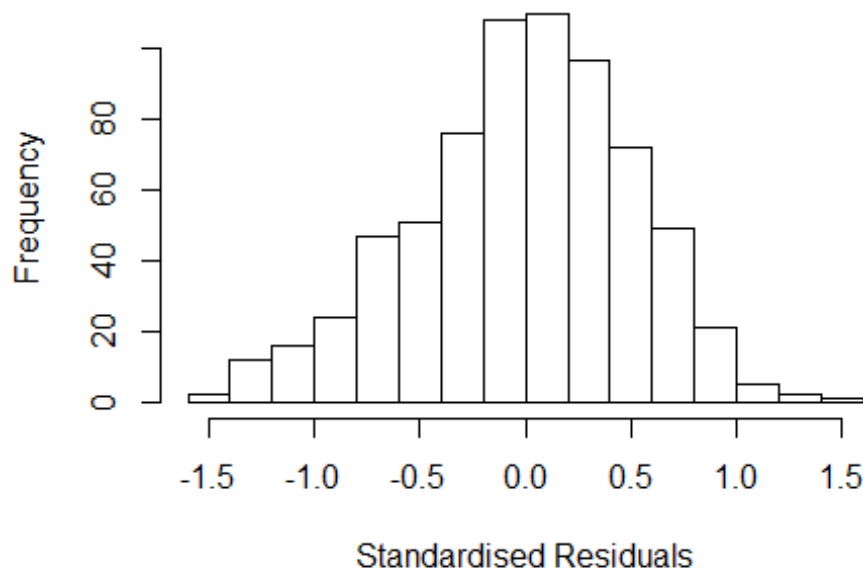
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 8e-04
```





```
## [1] "Female first author team size 2018 geometric mean: 2.42556498714542"
## [1] "Male first author team size 2018 geometric mean: 2.03920587632142"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 710, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.28123449861856"
## [1] "Male last author team size 2018 geometric mean: 2.11884067722873"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 550, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.609 1          1.269
## LastAuthorFemale  1.513 1          1.230
## UniqueAuthors     1.651 4          1.065
## Year               1.805 16         1.019
```

## Residuals from first and last author and team size



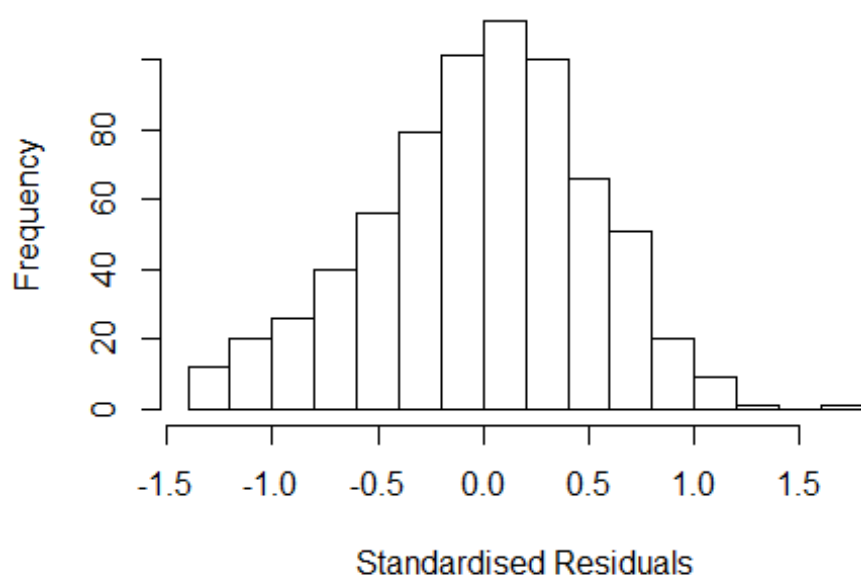
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4839 -0.3514  0.0224  0.3322  1.5070
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0362    0.0984   10.53  <2e-16 ***
## FirstAuthorFemale1  0.0246    0.0506    0.49  0.6274
## LastAuthorFemale1  0.0783    0.0513    1.53  0.1273
## UniqueAuthors2     0.0799    0.0463    1.73  0.0848 .
## UniqueAuthors3     0.1926    0.0661    2.92  0.0037 **
## UniqueAuthors4     0.0722    0.0874    0.83  0.4089
## UniqueAuthors5     0.3141    0.1388    2.26  0.0239 *
## Year1997           0.2720    0.1442    1.89  0.0596 .
## Year1998           0.1975    0.1601    1.23  0.2177
## Year1999           0.1155    0.1240    0.93  0.3518
```

```

## Year2000          0.1854      0.1232      1.51      0.1327
## Year2001          0.2551      0.1298      1.96      0.0498 *
## Year2002          0.2322      0.1277      1.82      0.0694 .
## Year2003          0.2183      0.1356      1.61      0.1077
## Year2004          0.2269      0.1610      1.41      0.1591
## Year2005          0.1138      0.1348      0.84      0.3988
## Year2006          0.0168      0.1277      0.13      0.8955
## Year2007          0.2040      0.1183      1.72      0.0853 .
## Year2008          0.0806      0.1069      0.75      0.4512
## Year2009          0.1356      0.1179      1.15      0.2502
## Year2010          0.0500      0.1162      0.43      0.6672
## Year2011          0.1733      0.1254      1.38      0.1674
## Year2012          0.0616      0.1112      0.55      0.5799
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0491, Adjusted R-squared:  0.0178
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 57 weights are ~= 1. The remaining 636 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.361  0.859  0.950  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.622 1      1.273
## LastAuthorFemale  1.555 1      1.247
## Year              1.155 16      1.005

```

## Residuals from first and last author



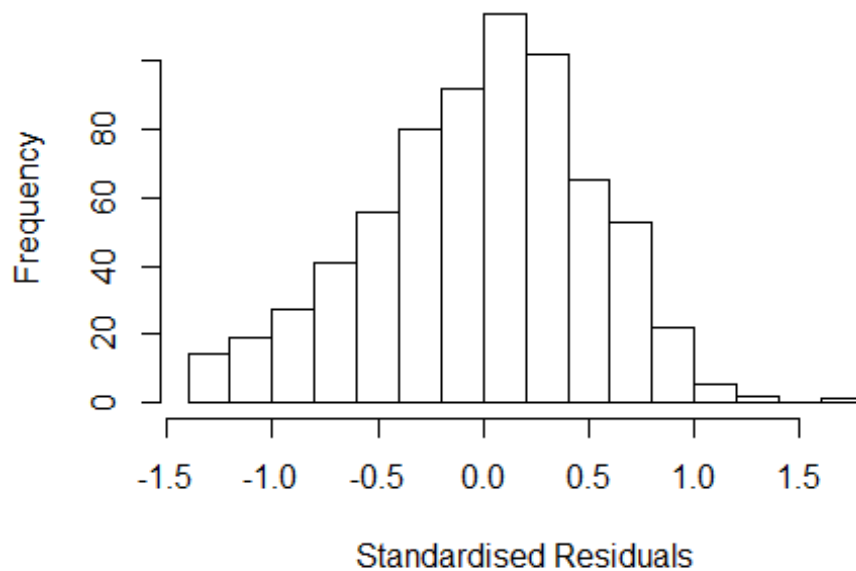
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3453 -0.3533 0.0246 0.3427 1.6336
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0742 0.0928 11.58 <2e-16 ***
## FirstAuthorFemale1 0.0404 0.0515 0.78 0.433
## LastAuthorFemale1 0.0707 0.0529 1.34 0.182
## Year1997 0.2767 0.1438 1.92 0.055 .
## Year1998 0.1973 0.1552 1.27 0.204
## Year1999 0.1237 0.1233 1.00 0.316
## Year2000 0.2137 0.1227 1.74 0.082 .
## Year2001 0.2711 0.1277 2.12 0.034 *
## Year2002 0.2602 0.1291 2.02 0.044 *
## Year2003 0.2239 0.1382 1.62 0.106
## Year2004 0.2515 0.1609 1.56 0.119
## Year2005 0.1161 0.1342 0.87 0.387
```

```

## Year2006          0.0363      0.1272      0.29      0.775
## Year2007          0.2385      0.1178      2.02      0.043 *
## Year2008          0.1156      0.1059      1.09      0.275
## Year2009          0.1664      0.1172      1.42      0.156
## Year2010          0.0780      0.1165      0.67      0.504
## Year2011          0.1982      0.1248      1.59      0.113
## Year2012          0.1029      0.1093      0.94      0.347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0288, Adjusted R-squared:  0.00289
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 51 weights are ~= 1. The remaining 642 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.296  0.860  0.953  0.905  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1      1.044
## Year              1.091 16      1.003

```

## Residuals from first author



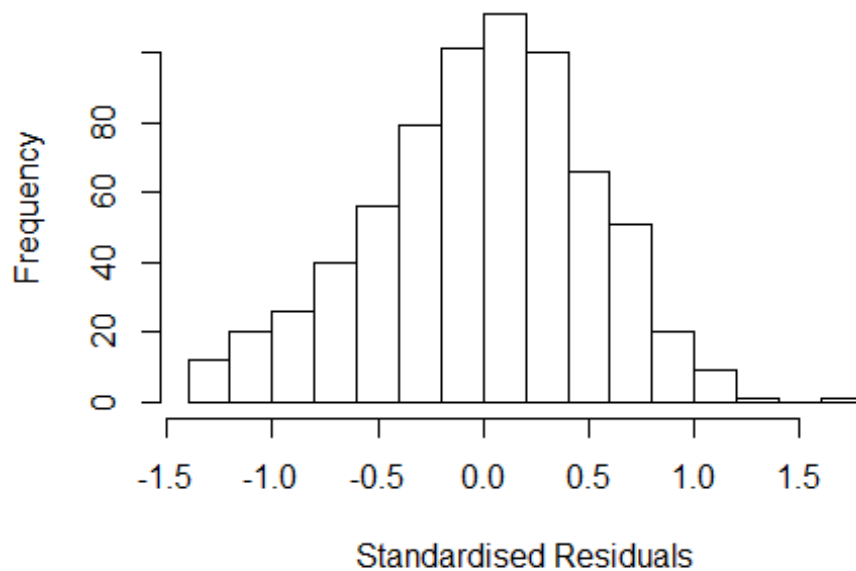
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3505 -0.3489 0.0265 0.3475 1.6268
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0870 0.0944 11.51 <2e-16 ***
## FirstAuthorFemale1 0.0756 0.0424 1.79 0.075 .
## Year1997 0.2698 0.1456 1.85 0.064 .
## Year1998 0.1931 0.1572 1.23 0.220
## Year1999 0.1156 0.1251 0.92 0.356
## Year2000 0.2099 0.1236 1.70 0.090 .
## Year2001 0.2635 0.1296 2.03 0.042 *
## Year2002 0.2542 0.1299 1.96 0.051 .
## Year2003 0.2196 0.1409 1.56 0.120
## Year2004 0.2565 0.1608 1.60 0.111
## Year2005 0.1044 0.1350 0.77 0.439
## Year2006 0.0283 0.1281 0.22 0.825
```

```

## Year2007          0.2372      0.1192      1.99      0.047 *
## Year2008          0.1164      0.1079      1.08      0.281
## Year2009          0.1645      0.1195      1.38      0.169
## Year2010          0.0714      0.1178      0.61      0.545
## Year2011          0.1983      0.1268      1.56      0.118
## Year2012          0.0933      0.1108      0.84      0.400
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0267, Adjusted R-squared:  0.00218
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 641 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.299  0.856  0.951  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.045 1      1.022
## Year              1.045 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3514 -0.3522 0.0269 0.3454 1.6316
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0773 0.0921 11.70 <2e-16 ***
## LastAuthorFemale1 0.0919 0.0434 2.12 0.035 *
## Year1997 0.2823 0.1440 1.96 0.050 .
## Year1998 0.1991 0.1552 1.28 0.200
## Year1999 0.1261 0.1227 1.03 0.304
## Year2000 0.2194 0.1218 1.80 0.072 .
## Year2001 0.2741 0.1274 2.15 0.032 *
## Year2002 0.2591 0.1286 2.01 0.044 *
## Year2003 0.2240 0.1368 1.64 0.102
## Year2004 0.2461 0.1603 1.54 0.125
## Year2005 0.1202 0.1336 0.90 0.369
## Year2006 0.0380 0.1271 0.30 0.765
```



```

## Year2007          0.2415      0.1175      2.06      0.040 *
## Year2008          0.1186      0.1056      1.12      0.262
## Year2009          0.1724      0.1166      1.48      0.140
## Year2010          0.0834      0.1166      0.72      0.475
## Year2011          0.2009      0.1241      1.62      0.106
## Year2012          0.1098      0.1090      1.01      0.314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.0281, Adjusted R-squared:  0.00365
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 640 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.296  0.860  0.952  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 693"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1801"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2005 2007 2009 2010 2011 2012
##    1    1    4    1    4    3
##
## 2005 2007 2009 2010 2011 2012
##    0    1    4    0    4    3
##
## 2005 2007 2009 2010 2011 2012
##    0    1    3    0    4    3

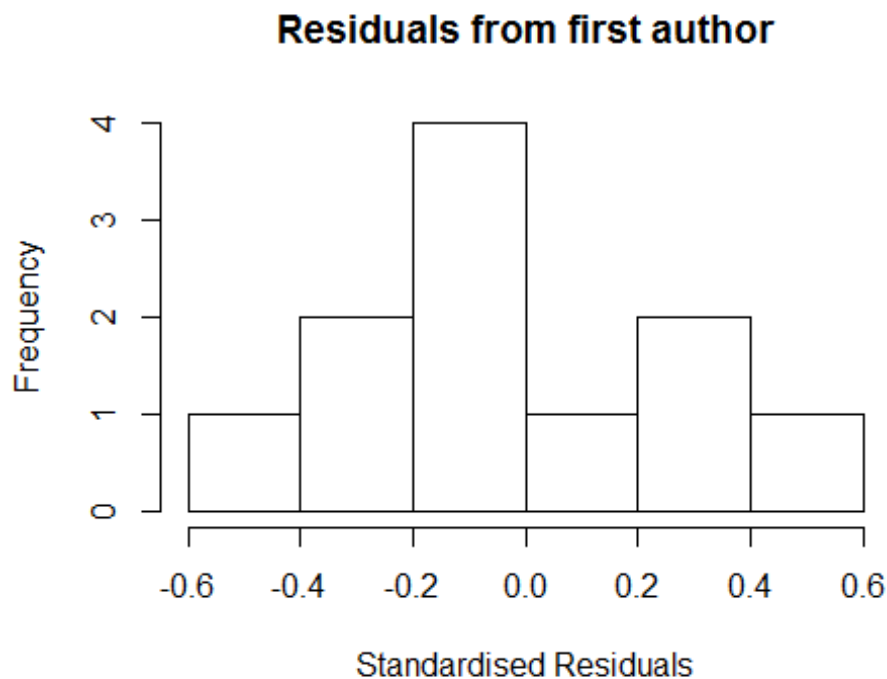
```

```
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.002e+15  1      5.479e+07
## Year              3.002e+15  3      3.798e+02

## [1] "Regression 4: Last author gender, Year as factors"

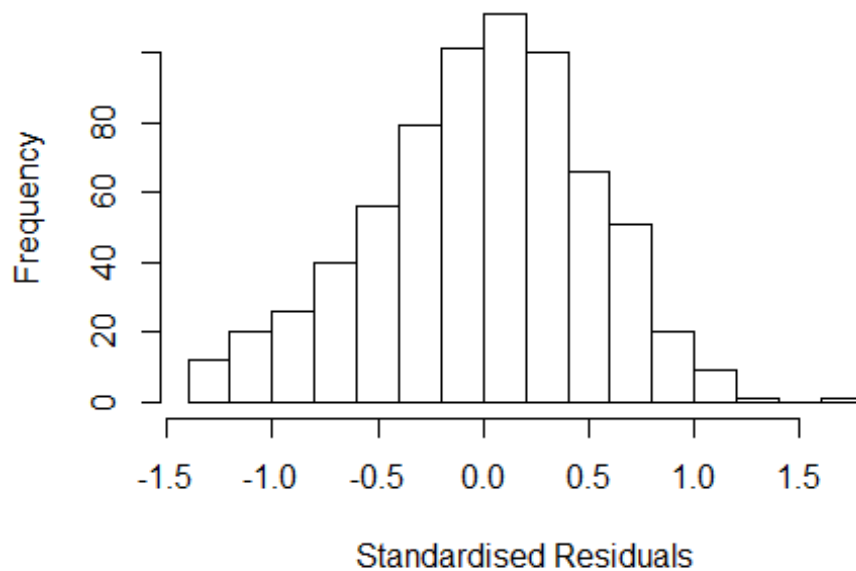
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```



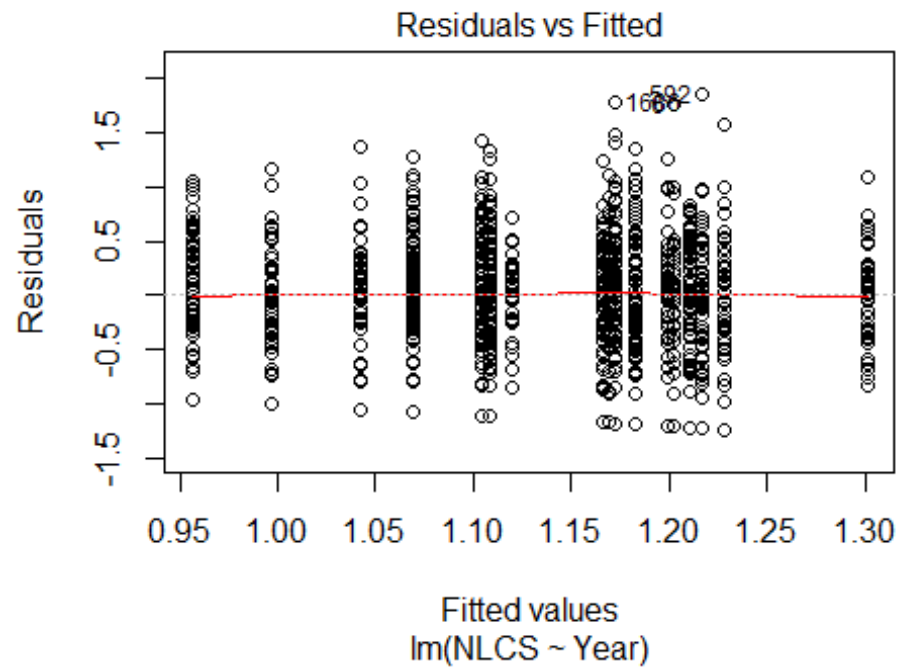
```
##
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale NaN  1      NaN
## Year             NaN  3      NaN
```

## Residuals from last author

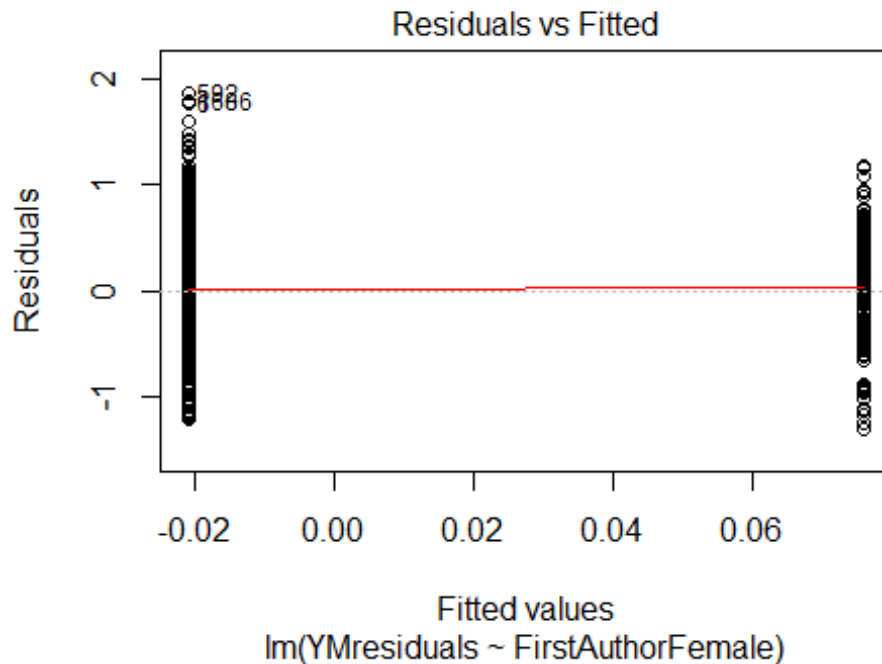


```
## [1] "Sample size for the above analysis: 11"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1802"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    61    69    85    98   120    94    63    50    69    64   135   145   124   127   133
## 2011 2012
##   101   112
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    39    43    51    56    85    56    44    30    47    41   102   103    88    95    91
## 2011 2012
##    68    77
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    35    34    51    56    80    54    42    24    41    38    90    90    77    88    81
## 2011 2012
##    60    71
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
```

```
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.02
```

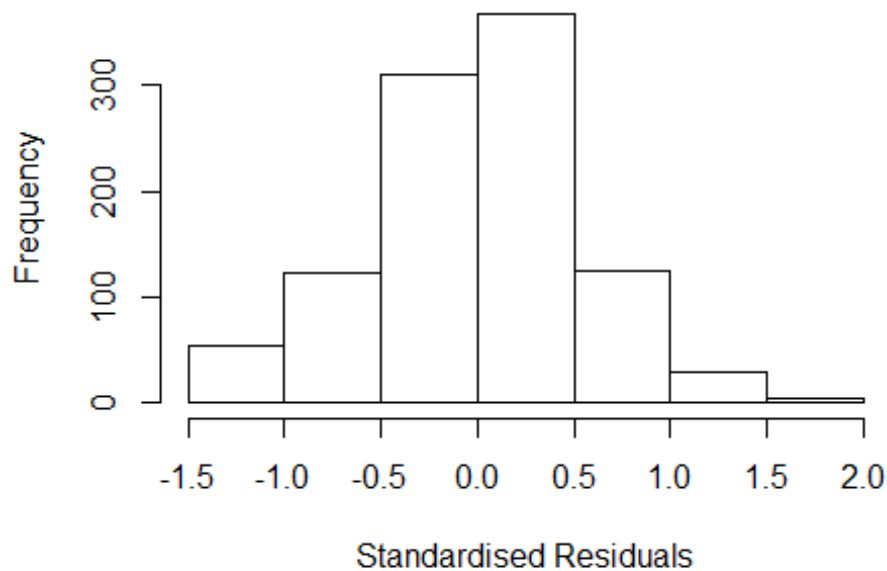


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.77675914518409"
## [1] "Male first author team size 2018 geometric mean: 2.02029056206252"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 690, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.13411509026839"
## [1] "Male last author team size 2018 geometric mean: 1.91272291238506"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.445 1          1.202
## LastAuthorFemale  1.477 1          1.215
## UniqueAuthors    1.336 4          1.037
## Year              1.428 16         1.011
```

## Residuals from first and last author and team size



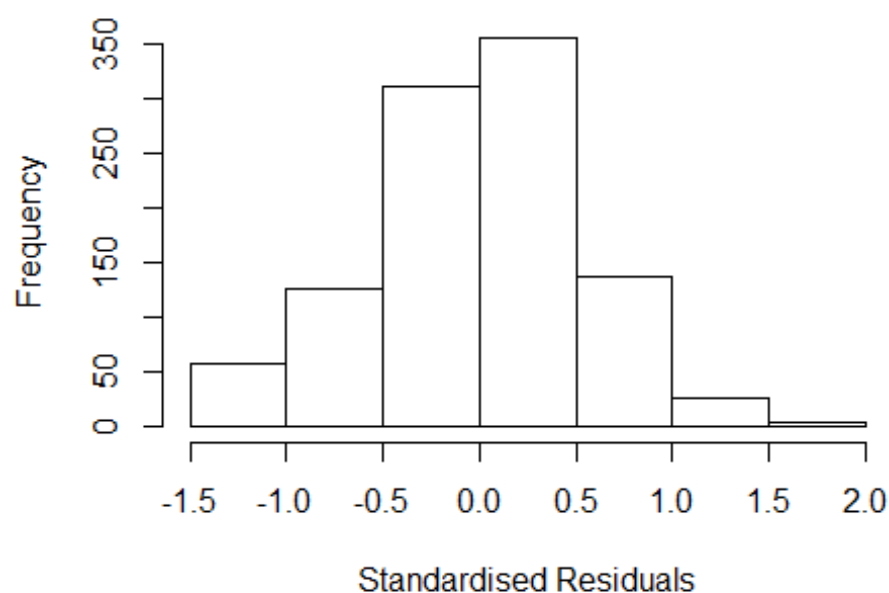
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.422 -0.331 0.019 0.343 1.861
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16e+00 7.31e-02 15.82 <2e-16 ***
## FirstAuthorFemale1 1.14e-01 4.59e-02 2.48 0.013 *
## LastAuthorFemale1 5.44e-03 4.78e-02 0.11 0.909
## UniqueAuthors2 -7.63e-02 4.16e-02 -1.84 0.067 .
## UniqueAuthors3 -5.08e-05 5.06e-02 0.00 0.999
## UniqueAuthors4 1.12e-01 7.61e-02 1.47 0.143
## UniqueAuthors5 -2.51e-01 1.08e-01 -2.34 0.020 *
## Year1997 7.13e-02 1.19e-01 0.60 0.550
## Year1998 -1.60e-01 9.82e-02 -1.63 0.103
## Year1999 -1.19e-01 9.62e-02 -1.23 0.218
```

```

## Year2000      -1.99e-01  9.11e-02  -2.18  0.029 *
## Year2001      8.26e-02  1.06e-01   0.78  0.435
## Year2002      3.52e-02  1.10e-01   0.32  0.749
## Year2003      2.92e-02  1.06e-01   0.28  0.783
## Year2004      5.07e-02  1.17e-01   0.43  0.666
## Year2005      1.35e-01  1.01e-01   1.34  0.180
## Year2006     -5.33e-02  9.40e-02  -0.57  0.571
## Year2007     -6.96e-02  9.65e-02  -0.72  0.471
## Year2008     -8.49e-02  1.03e-01  -0.83  0.409
## Year2009      2.48e-02  1.03e-01   0.24  0.809
## Year2010     -1.48e-02  1.08e-01  -0.14  0.891
## Year2011     -1.37e-02  1.01e-01  -0.14  0.893
## Year2012      5.24e-02  9.46e-02   0.55  0.579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.0441, Adjusted R-squared:  0.0229
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.155  0.863  0.949  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.471 1 1.213
## LastAuthorFemale 1.503 1 1.226
## Year 1.096 16 1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2970 -0.3311 0.0268 0.3474 1.8437
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14626 0.07490 15.30 <2e-16 ***
## FirstAuthorFemale1 0.11000 0.04688 2.35 0.019 *
## LastAuthorFemale1 0.00256 0.04894 0.05 0.958
## Year1997 0.05972 0.12086 0.49 0.621
## Year1998 -0.17831 0.09814 -1.82 0.070 .
## Year1999 -0.12424 0.09689 -1.28 0.200
## Year2000 -0.21222 0.09191 -2.31 0.021 *
## Year2001 0.06901 0.10584 0.65 0.515
## Year2002 0.02080 0.10928 0.19 0.849
## Year2003 -0.00691 0.10239 -0.07 0.946
## Year2004 0.03815 0.11639 0.33 0.743
## Year2005 0.12932 0.10019 1.29 0.197
```

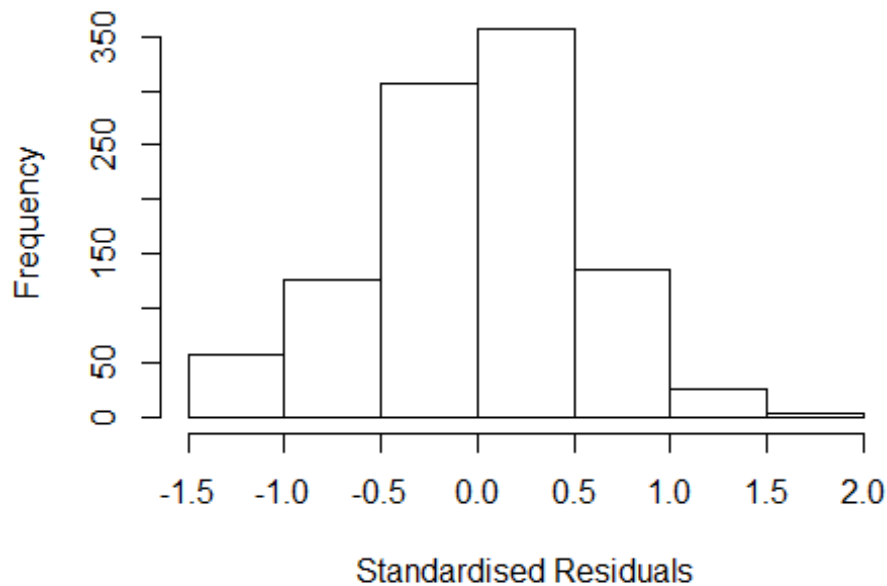


```

## Year2006      -0.07172    0.09415   -0.76    0.446
## Year2007      -0.09417    0.09526   -0.99    0.323
## Year2008      -0.08742    0.10337   -0.85    0.398
## Year2009       0.02035    0.10270    0.20    0.843
## Year2010      -0.02786    0.10676   -0.26    0.794
## Year2011      -0.01081    0.10095   -0.11    0.915
## Year2012       0.04671    0.09422    0.50    0.620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.034, Adjusted R-squared:  0.0165
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 927 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.171  0.865   0.950   0.900   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"    "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1      1.017
## Year              1.033 16      1.001

```

## Residuals from first author



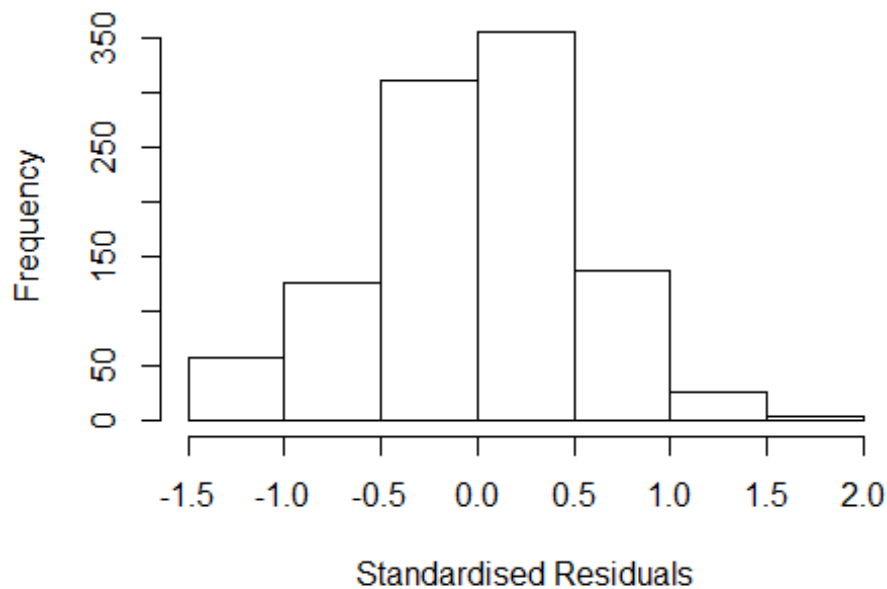
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2959 -0.3314 0.0261 0.3470 1.8437
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1465 0.0748 15.32 <2e-16 ***
## FirstAuthorFemale1 0.1112 0.0394 2.82 0.0049 **
## Year1997 0.0597 0.1209 0.49 0.6217
## Year1998 -0.1783 0.0981 -1.82 0.0696 .
## Year1999 -0.1243 0.0969 -1.28 0.1999
## Year2000 -0.2122 0.0919 -2.31 0.0211 *
## Year2001 0.0688 0.1058 0.65 0.5154
## Year2002 0.0210 0.1092 0.19 0.8476
## Year2003 -0.0068 0.1023 -0.07 0.9471
## Year2004 0.0382 0.1164 0.33 0.7426
## Year2005 0.1294 0.1002 1.29 0.1968
## Year2006 -0.0716 0.0941 -0.76 0.4469
```

```

## Year2007          -0.0942      0.0951   -0.99   0.3219
## Year2008          -0.0874      0.1033   -0.85   0.3979
## Year2009           0.0205      0.1026    0.20   0.8417
## Year2010          -0.0278      0.1068   -0.26   0.7946
## Year2011          -0.0107      0.1009   -0.11   0.9157
## Year2012           0.0469      0.0942    0.50   0.6189
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.0341, Adjusted R-squared:  0.0175
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 927 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.170  0.865   0.949   0.900   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.063 1          1.031
## Year            1.063 16          1.002

```

## Residuals from last author



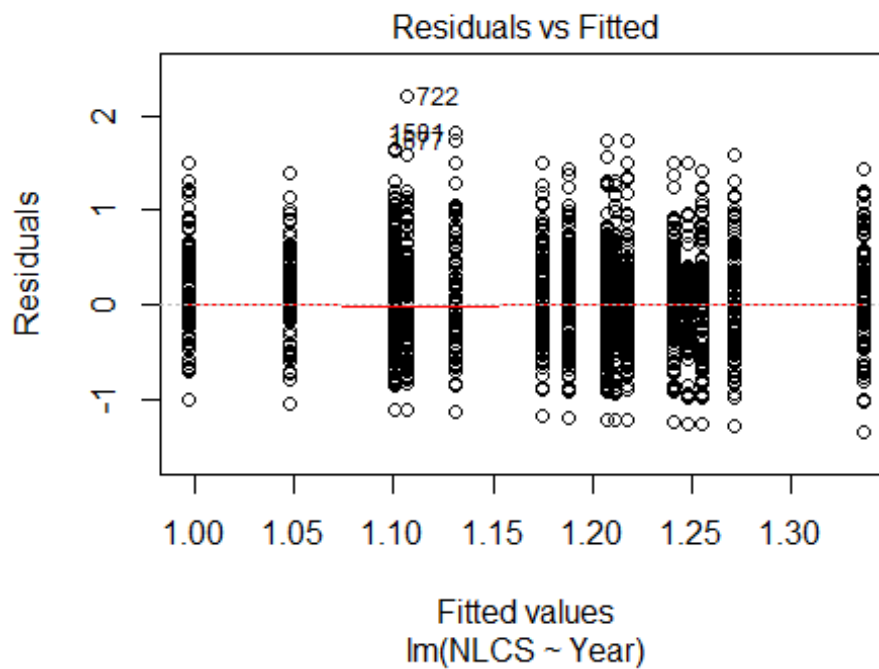
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2535 -0.3279 0.0261 0.3397 1.8321
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16642 0.07544 15.46 <2e-16 ***
## LastAuthorFemale1 0.05768 0.04181 1.38 0.168
## Year1997 0.05053 0.12087 0.42 0.676
## Year1998 -0.18817 0.09972 -1.89 0.059 .
## Year1999 -0.12538 0.09861 -1.27 0.204
## Year2000 -0.21681 0.09314 -2.33 0.020 *
## Year2001 0.06050 0.10672 0.57 0.571
## Year2002 0.01413 0.10966 0.13 0.897
## Year2003 -0.01951 0.10547 -0.18 0.853
## Year2004 0.02935 0.11679 0.25 0.802
## Year2005 0.12881 0.10087 1.28 0.202
## Year2006 -0.07958 0.09528 -0.84 0.404
```

```

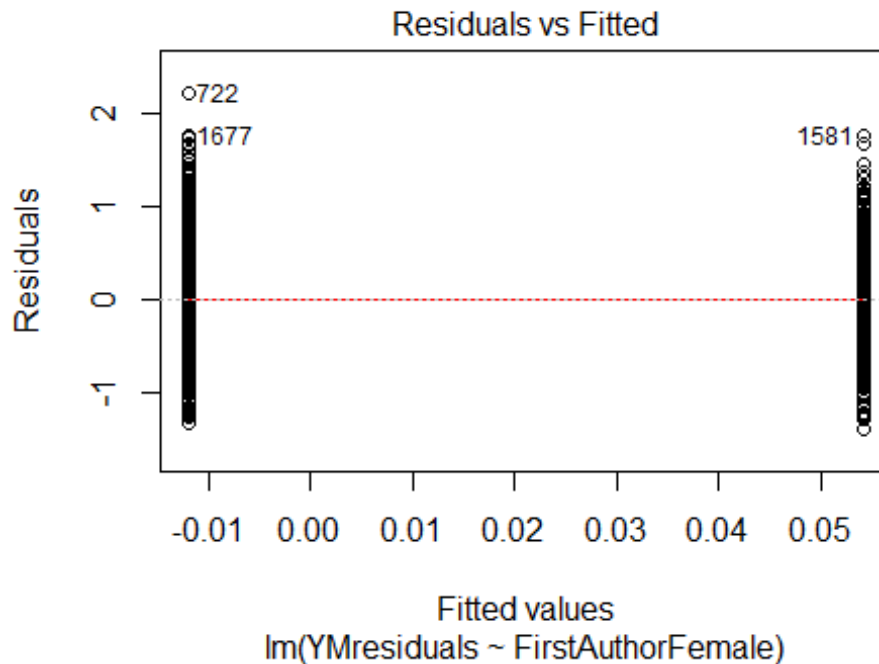
## Year2007          -0.10345      0.09664      -1.07      0.285
## Year2008          -0.08829      0.10543      -0.84      0.403
## Year2009           0.00886      0.10305       0.09      0.931
## Year2010          -0.02946      0.10809      -0.27      0.785
## Year2011          -0.01525      0.10255      -0.15      0.882
## Year2012           0.03982      0.09467       0.42      0.674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.0284, Adjusted R-squared:  0.0118
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 915 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.178  0.859  0.949  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1012"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1803"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 250 234 258 253 312 224 229 191 178 189 231 278 244 239 216
## 2011 2012
## 199 210
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 135 110 116 124 161 96 118 92 88 102 132 164 138 128 129
## 2011 2012

```

```
## 116 134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 129 102 110 120 150 91 107 87 82 93 117 145 120 114 105
## 2011 2012
## 96 107
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.06
```

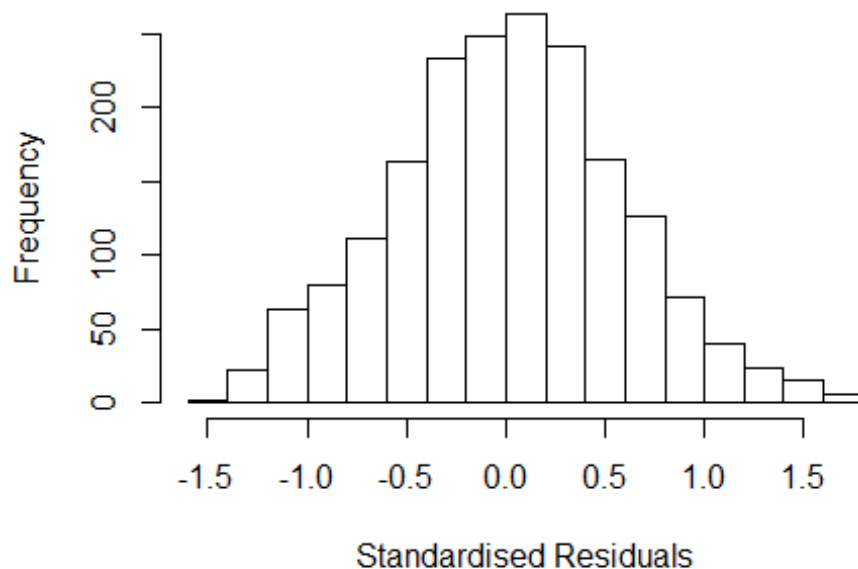


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.06, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 1.91593505702146"
## [1] "Male first author team size 2018 geometric mean: 2.28485744526616"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.10809499794134"
## [1] "Male last author team size 2018 geometric mean: 2.22831102737029"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.197 1      1.094
## LastAuthorFemale  1.192 1      1.092
## UniqueAuthors     1.224 4      1.026
## Year              1.285 16     1.008
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4132 -0.3830 0.0121 0.3776 1.7785
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9130 0.0541 16.87 < 2e-16 ***
## FirstAuthorFemale1 0.0455 0.0384 1.19 0.23572
## LastAuthorFemale1 0.0341 0.0390 0.88 0.38146
## UniqueAuthors2 0.1343 0.0330 4.07 5.0e-05 ***
## UniqueAuthors3 0.2123 0.0394 5.39 7.9e-08 ***
## UniqueAuthors4 0.3031 0.0599 5.06 4.5e-07 ***
## UniqueAuthors5 0.1120 0.0618 1.81 0.07042 .
## Year1997 0.0722 0.0823 0.88 0.38020
## Year1998 0.0647 0.0796 0.81 0.41601
## Year1999 0.0925 0.0831 1.11 0.26577
```

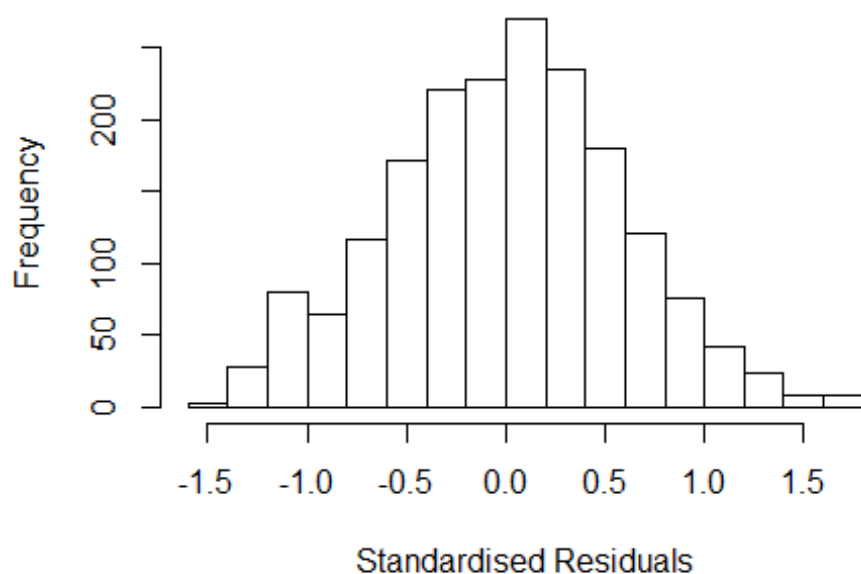


```

## Year2000          0.0415      0.0744      0.56  0.57720
## Year2001          0.1055      0.0940      1.12  0.26160
## Year2002          0.2450      0.0793      3.09  0.00204 **
## Year2003          0.1795      0.0749      2.40  0.01663 *
## Year2004          0.2312      0.0736      3.14  0.00171 **
## Year2005          0.2055      0.0824      2.49  0.01271 *
## Year2006          0.1378      0.0761      1.81  0.07037 .
## Year2007          0.1304      0.0710      1.84  0.06632 .
## Year2008          0.1918      0.0777      2.47  0.01366 *
## Year2009          0.1602      0.0753      2.13  0.03343 *
## Year2010          0.1448      0.0764      1.90  0.05818 .
## Year2011          0.3204      0.0901      3.55  0.00039 ***
## Year2012          0.1785      0.0745      2.40  0.01671 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.05,   Adjusted R-squared:  0.0387
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1717 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.315  0.866  0.953   0.906   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.188 1      1.090
## LastAuthorFemale  1.178 1      1.085
## Year              1.063 16      1.002

```

## Residuals from first and last author



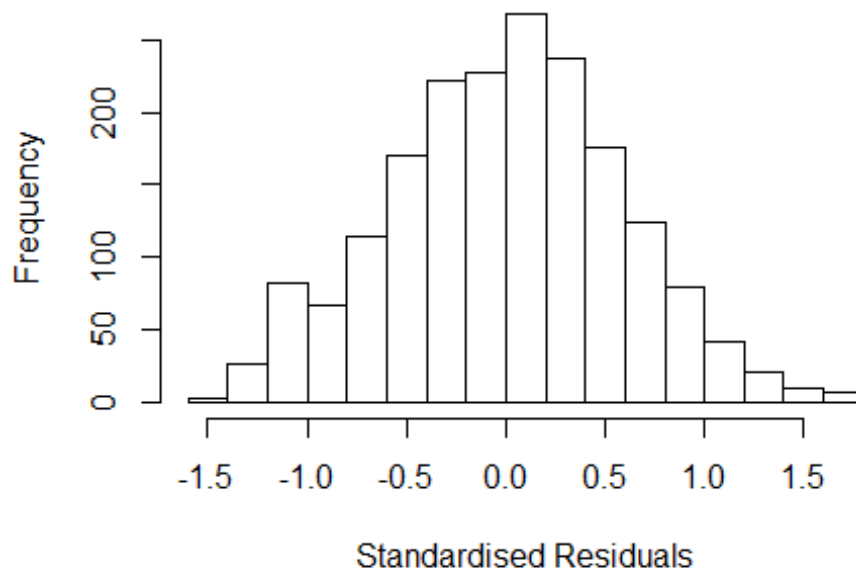
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4461 -0.3934 0.0236 0.3937 1.7880
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9684 0.0545 17.76 < 2e-16 ***
## FirstAuthorFemale1 0.0574 0.0387 1.48 0.13796
## LastAuthorFemale1 0.0362 0.0397 0.91 0.36193
## Year1997 0.0829 0.0829 1.00 0.31732
## Year1998 0.0723 0.0808 0.89 0.37138
## Year1999 0.1106 0.0841 1.32 0.18848
## Year2000 0.0950 0.0755 1.26 0.20850
## Year2001 0.1302 0.0941 1.38 0.16651
## Year2002 0.2726 0.0811 3.36 0.00079 ***
## Year2003 0.2199 0.0763 2.88 0.00399 **
## Year2004 0.2652 0.0743 3.57 0.00037 ***
## Year2005 0.2540 0.0833 3.05 0.00234 **
```

```

## Year2006          0.1909      0.0777      2.46  0.01407 *
## Year2007          0.1784      0.0725      2.46  0.01398 *
## Year2008          0.2371      0.0784      3.02  0.00254 **
## Year2009          0.2011      0.0758      2.65  0.00804 **
## Year2010          0.2069      0.0759      2.72  0.00650 **
## Year2011          0.3841      0.0893      4.30  1.8e-05 ***
## Year2012          0.2577      0.0736      3.50  0.00047 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.58
## Multiple R-squared:  0.0268, Adjusted R-squared:  0.0174
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 151 weights are ~= 1. The remaining 1724 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.321  0.872  0.952  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1      1.019
## Year              1.038 16      1.001

```

## Residuals from first author



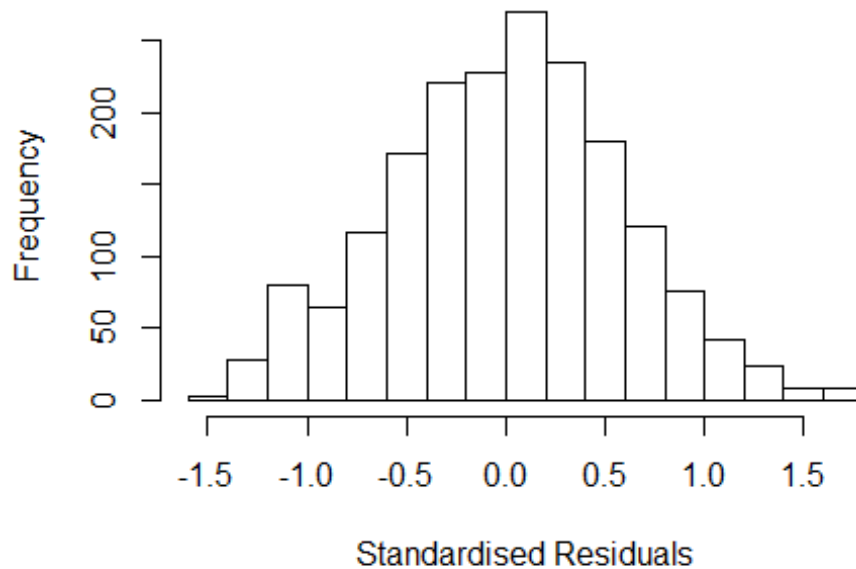
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4283 -0.3928 0.0216 0.3918 1.7699
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9720 0.0545 17.83 < 2e-16 ***
## FirstAuthorFemale1 0.0727 0.0362 2.01 0.04493 *
## Year1997 0.0813 0.0829 0.98 0.32708
## Year1998 0.0722 0.0810 0.89 0.37279
## Year1999 0.1098 0.0840 1.31 0.19105
## Year2000 0.0945 0.0756 1.25 0.21131
## Year2001 0.1294 0.0941 1.37 0.16931
## Year2002 0.2732 0.0809 3.38 0.00075 ***
## Year2003 0.2202 0.0761 2.89 0.00386 **
## Year2004 0.2676 0.0743 3.60 0.00032 ***
## Year2005 0.2566 0.0831 3.09 0.00204 **
## Year2006 0.1917 0.0778 2.46 0.01380 *
```

```

## Year2007          0.1774      0.0725      2.45  0.01448 *
## Year2008          0.2374      0.0784      3.03  0.00250 **
## Year2009          0.2013      0.0758      2.66  0.00798 **
## Year2010          0.2050      0.0758      2.70  0.00693 **
## Year2011          0.3837      0.0893      4.29  1.8e-05 ***
## Year2012          0.2560      0.0737      3.47  0.00053 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.0263, Adjusted R-squared:  0.0174
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 159 weights are ~= 1. The remaining 1716 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.873  0.951  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.031 1      1.016
## Year              1.031 16      1.001

```

## Residuals from last author



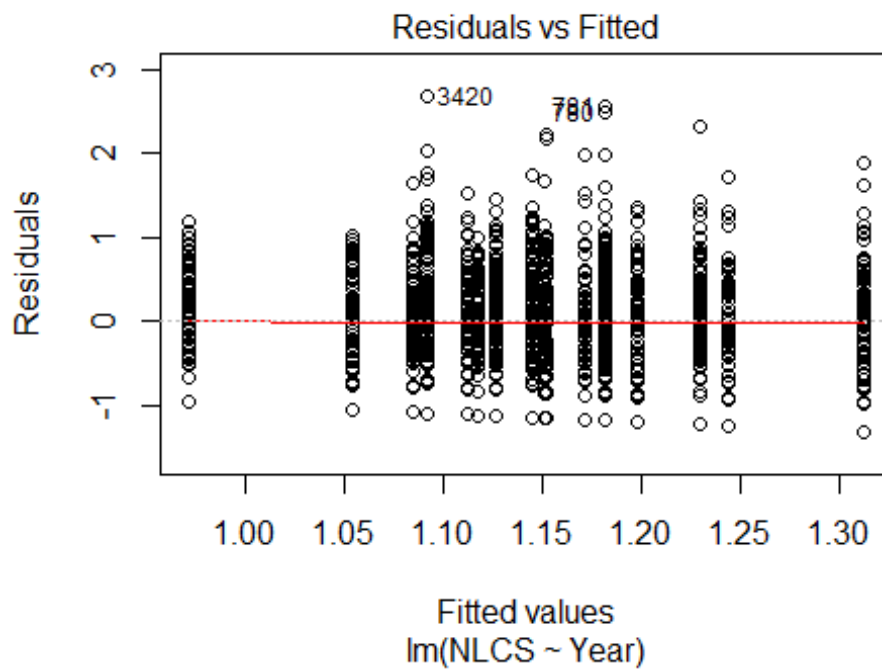
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4223 -0.4000 0.0182 0.3873 1.8385
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9738 0.0545 17.87 < 2e-16 ***
## LastAuthorFemale1 0.0613 0.0371 1.65 0.09826 .
## Year1997 0.0805 0.0830 0.97 0.33235
## Year1998 0.0711 0.0807 0.88 0.37871
## Year1999 0.1126 0.0839 1.34 0.17946
## Year2000 0.0950 0.0754 1.26 0.20803
## Year2001 0.1317 0.0943 1.40 0.16250
## Year2002 0.2734 0.0809 3.38 0.00074 ***
## Year2003 0.2194 0.0765 2.87 0.00418 **
## Year2004 0.2612 0.0743 3.52 0.00045 ***
## Year2005 0.2514 0.0836 3.01 0.00268 **
## Year2006 0.1938 0.0775 2.50 0.01251 *
```

```

## Year2007          0.1817      0.0726      2.50  0.01236 *
## Year2008          0.2368      0.0788      3.00  0.00270 **
## Year2009          0.2010      0.0759      2.65  0.00817 **
## Year2010          0.2086      0.0762      2.74  0.00625 **
## Year2011          0.3872      0.0893      4.33  1.5e-05 ***
## Year2012          0.2621      0.0739      3.55  0.00040 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.58
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.0168
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 155 weights are ~= 1. The remaining 1720 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.293  0.872  0.951  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1875"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1804"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 117 135 131 139 140 140 145 106 145 143 197 206 206 184
## 2011 2012
## 207 199
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 72 60 78 66 71 85 88 96 69 106 87 120 127 125 128
## 2011 2012

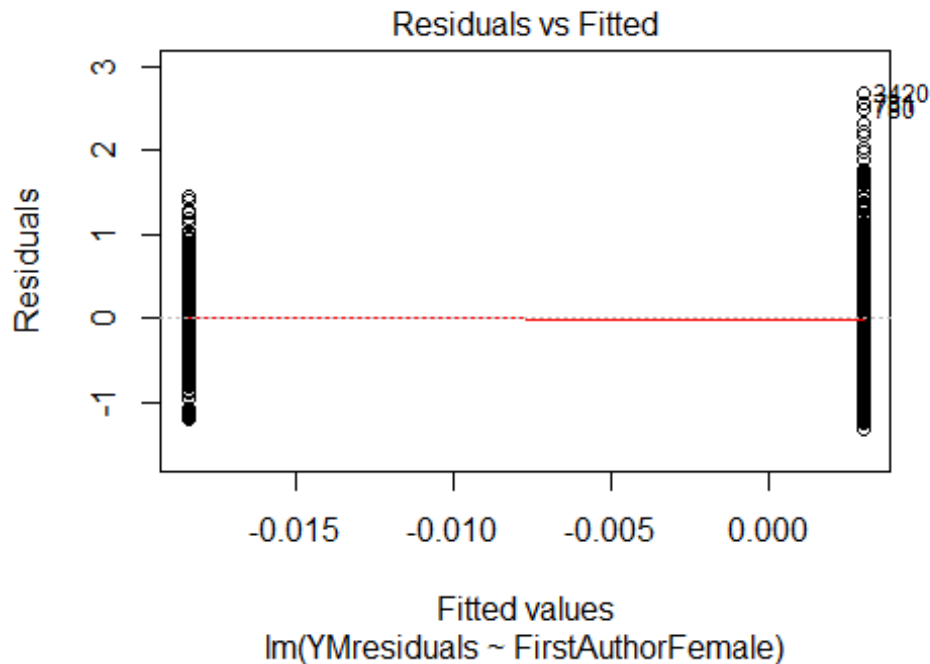
```

```
## 146 142
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 53 72 65 63 77 79 88 67 98 78 107 119 110 120
## 2011 2012
## 136 127
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 2e-04
```



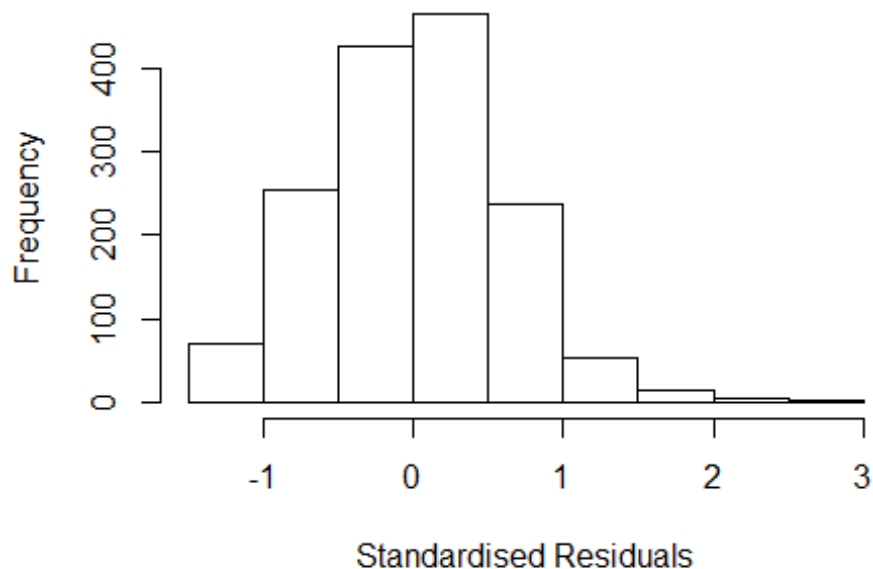
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07
```





```
## [1] "Female first author team size 2018 geometric mean: 2.10393133964896"
## [1] "Male first author team size 2018 geometric mean: 1.8324267390422"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.80290806703324"
## [1] "Male last author team size 2018 geometric mean: 1.90406123786058"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.260 1          1.122
## LastAuthorFemale  1.266 1          1.125
## UniqueAuthors    1.176 4          1.021
## Year              1.239 16         1.007
```

## Residuals from first and last author and team size



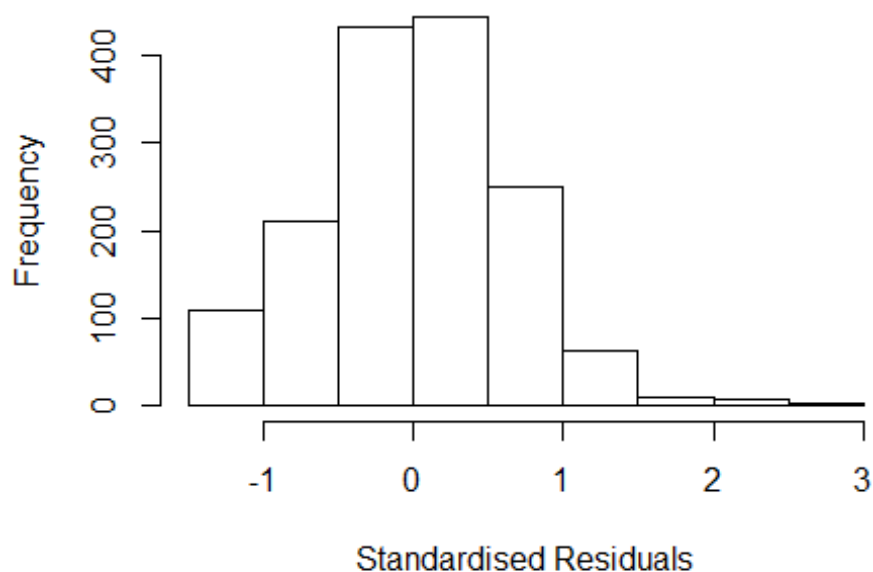
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3161 77749249761 3.553 2010    1712     3     2.508
## 3420 78650862532 3.769 2011    1804     2     2.898
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4062 -0.4114  0.0151  0.4231  2.8980
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.2075     0.0832   14.51 < 2e-16 ***
## FirstAuthorFemale1 -0.0654     0.0477   -1.37  0.17084
## LastAuthorFemale1  0.0695     0.0499    1.39  0.16356
## UniqueAuthors2     0.1987     0.0367    5.42  7.1e-08 ***
## UniqueAuthors3     0.3119     0.0477    6.54  8.6e-11 ***
## UniqueAuthors4     0.3868     0.0799    4.84  1.4e-06 ***
## UniqueAuthors5     0.6320     0.1321    4.78  1.9e-06 ***
## Year1997        -0.0440     0.1149   -0.38  0.70162
## Year1998        -0.0783     0.1167   -0.67  0.50267
```

```

## Year1999          -0.1863      0.1217    -1.53   0.12601
## Year2000          -0.2312      0.1143    -2.02   0.04333 *
## Year2001          -0.2487      0.1100    -2.26   0.02397 *
## Year2002          -0.2186      0.1042    -2.10   0.03613 *
## Year2003          -0.2123      0.0960    -2.21   0.02711 *
## Year2004          -0.2352      0.1103    -2.13   0.03318 *
## Year2005          -0.3563      0.1033    -3.45   0.00058 ***
## Year2006          -0.2919      0.1040    -2.81   0.00507 **
## Year2007          -0.2716      0.0989    -2.75   0.00612 **
## Year2008          -0.2123      0.0981    -2.16   0.03069 *
## Year2009          -0.1975      0.1028    -1.92   0.05487 .
## Year2010          -0.1628      0.1018    -1.60   0.11001
## Year2011          -0.3365      0.1016    -3.31   0.00094 ***
## Year2012          -0.2301      0.1068    -2.15   0.03133 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.612
## Multiple R-squared:  0.0651, Adjusted R-squared:  0.0514
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 1367 is an outlier with |weight| = 0 ( < 6.5e-05);
## 154 weights are ~= 1. The remaining 1373 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0547 0.8750 0.9460 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.54e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.239 1 1.113
## LastAuthorFemale 1.258 1 1.122
## Year 1.073 16 1.002

```

## Residuals from first and last author



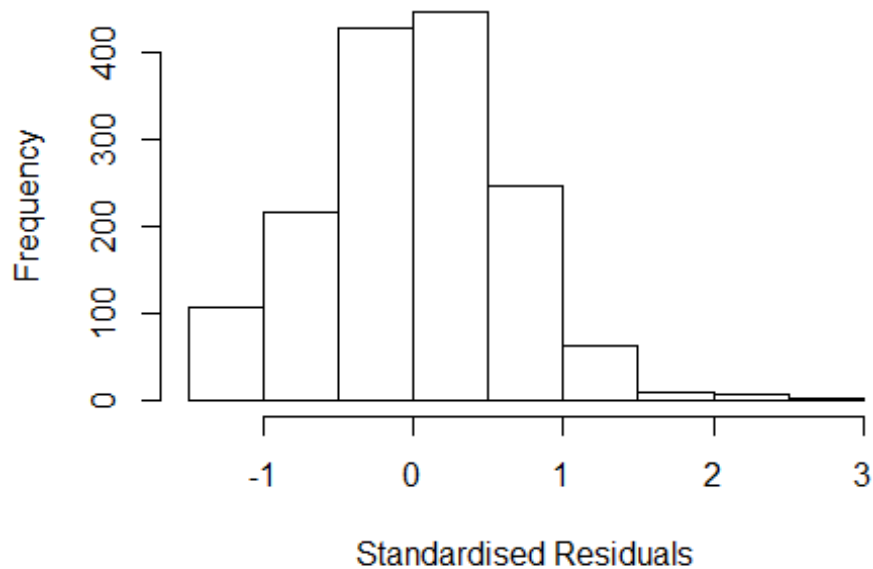
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 780   0001460136 3.678 2000    1703     4    2.595
## 781   0006407254 3.743 2000    1703     4    2.660
## 3420 78650862532 3.769 2011    1804     2    2.735
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.292 -0.425  0.013  0.434  2.735
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2920     0.0855   15.11  <2e-16 ***
## FirstAuthorFemale1 -0.0310     0.0487    -0.64  0.5249
## LastAuthorFemale1  0.0755     0.0531     1.42  0.1553
## Year1997          -0.0688     0.1182    -0.58  0.5607
## Year1998          -0.0944     0.1220    -0.77  0.4390
## Year1999          -0.1737     0.1259    -1.38  0.1678
## Year2000          -0.2088     0.1167    -1.79  0.0739 .
## Year2001          -0.2414     0.1113    -2.17  0.0302 *
## Year2002          -0.1855     0.1068    -1.74  0.0825 .
## Year2003          -0.1906     0.0985    -1.93  0.0532 .
```

```

## Year2004          -0.1767      0.1103    -1.60    0.1093
## Year2005          -0.3277      0.1052    -3.11    0.0019 **
## Year2006          -0.2387      0.1065    -2.24    0.0251 *
## Year2007          -0.2202      0.1012    -2.18    0.0296 *
## Year2008          -0.1656      0.1016    -1.63    0.1031
## Year2009          -0.1550      0.1072    -1.45    0.1485
## Year2010          -0.0897      0.1060    -0.85    0.3973
## Year2011          -0.2576      0.1052    -2.45    0.0145 *
## Year2012          -0.1736      0.1099    -1.58    0.1143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.00366
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1384 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0177 0.8700 0.9490 0.9070 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1      1.016
## Year              1.031 16      1.001

```

## Residuals from first author



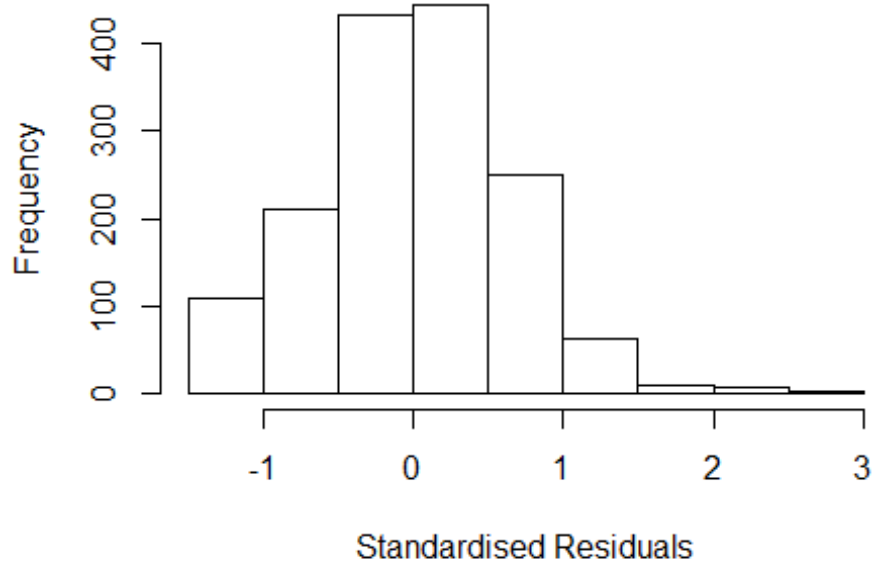
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 780   0001460136 3.678 2000    1703     4    2.595
## 781   0006407254 3.743 2000    1703     4    2.660
## 3420 78650862532 3.769 2011    1804     2    2.735
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.30004 -0.42837  0.00788  0.42927  2.73169
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.30004    0.08533   15.23  <2e-16 ***
## FirstAuthorFemale1 -0.00252    0.04394   -0.06  0.9543
## Year1997        -0.06902    0.11797   -0.59  0.5586
## Year1998        -0.09451    0.12227   -0.77  0.4397
## Year1999        -0.18132    0.12552   -1.44  0.1488
## Year2000        -0.20920    0.11639   -1.80  0.0725 .
## Year2001        -0.24658    0.11106   -2.22  0.0265 *
## Year2002        -0.18830    0.10656   -1.77  0.0774 .
## Year2003        -0.19480    0.09818   -1.98  0.0474 *
## Year2004        -0.18046    0.11006   -1.64  0.1013
```

```

## Year2005      -0.33161    0.10524   -3.15    0.0017 **
## Year2006      -0.24616    0.10660   -2.31    0.0211 *
## Year2007      -0.21889    0.10097   -2.17    0.0303 *
## Year2008      -0.16805    0.10144   -1.66    0.0978 .
## Year2009      -0.15792    0.10732   -1.47    0.1414
## Year2010      -0.09137    0.10598   -0.86    0.3888
## Year2011      -0.26273    0.10512   -2.50    0.0125 *
## Year2012      -0.17611    0.10954   -1.61    0.1081
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0141, Adjusted R-squared:  0.00302
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1392 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0182 0.8700 0.9500 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.046 1          1.023
## Year            1.046 16          1.001

```

## Residuals from last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 780   0001460136 3.678 2000    1703     4    2.595
## 781   0006407254 3.743 2000    1703     4    2.660
## 3420  78650862532 3.769 2011    1804     2    2.735
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2902 -0.4229  0.0122  0.4326  2.7385
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2902     0.0854   15.10 <2e-16 ***
## LastAuthorFemale1  0.0617     0.0487    1.27  0.2055
## Year1997        -0.0709     0.1181   -0.60  0.5483
## Year1998        -0.0953     0.1223   -0.78  0.4359
## Year1999        -0.1740     0.1259   -1.38  0.1672
## Year2000        -0.2108     0.1166   -1.81  0.0708 .
## Year2001        -0.2422     0.1112   -2.18  0.0297 *
## Year2002        -0.1868     0.1069   -1.75  0.0807 .
## Year2003        -0.1923     0.0987   -1.95  0.0516 .
## Year2004        -0.1774     0.1102   -1.61  0.1076
```

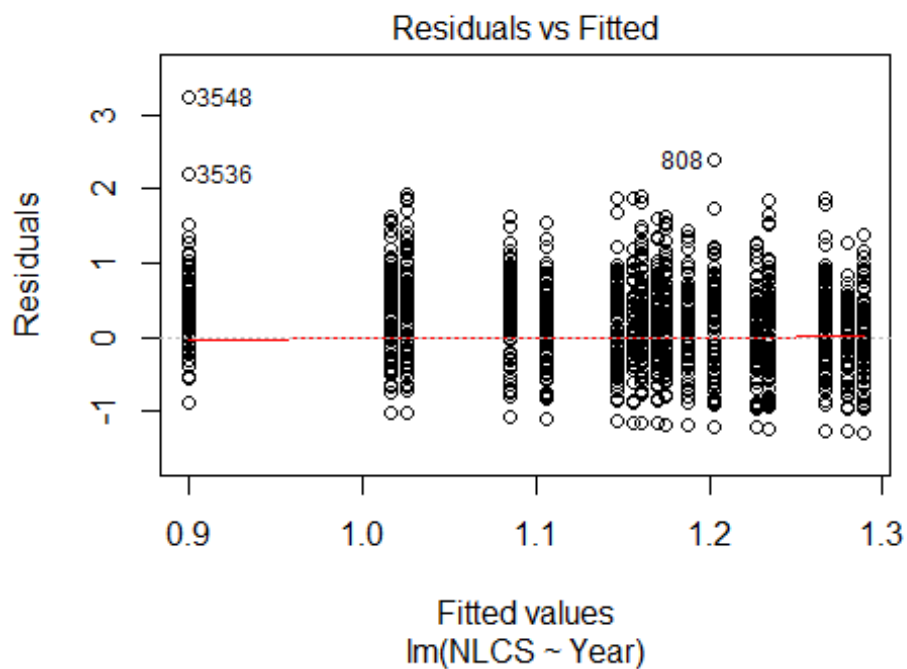


```

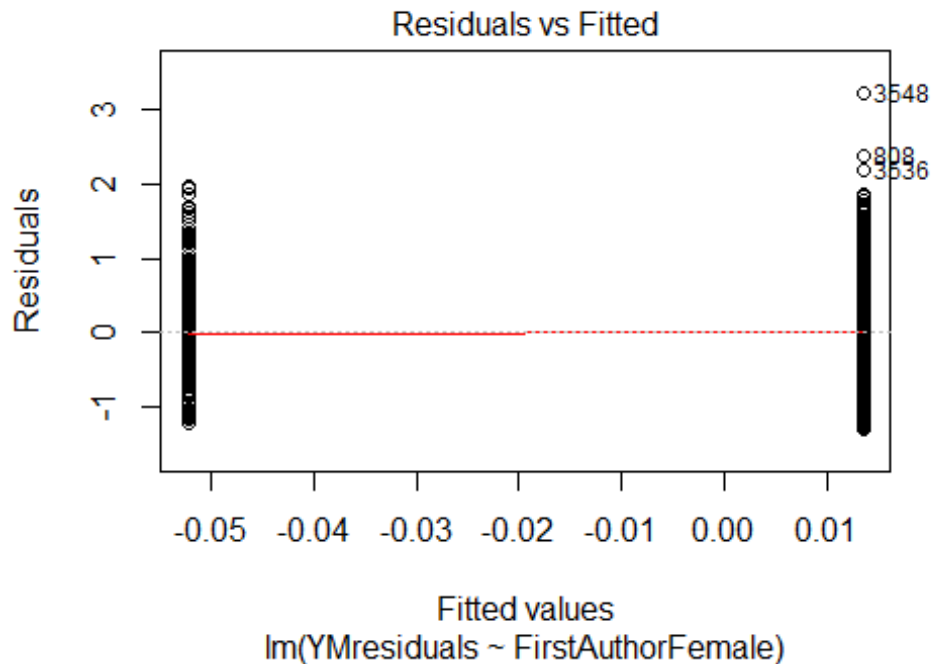
## Year2005          -0.3272      0.1053    -3.11    0.0019 **
## Year2006          -0.2399      0.1065    -2.25    0.0245 *
## Year2007          -0.2213      0.1013    -2.18    0.0291 *
## Year2008          -0.1663      0.1015    -1.64    0.1017
## Year2009          -0.1558      0.1072    -1.45    0.1465
## Year2010          -0.0918      0.1061    -0.87    0.3870
## Year2011          -0.2597      0.1052    -2.47    0.0136 *
## Year2012          -0.1758      0.1100    -1.60    0.1100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.00405
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1391 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0174 0.8690 0.9490 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1528"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1900"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 326 325 235 199 217 238 153 172 181 135 160 166 181 204 228
## 2011 2012
## 181 173
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 127 104 97 82 62 57 88 102 88 81 98 102 121 131 145
## 2011 2012
## 110 120
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 121 99 91 75 57 55 79 92 83 67 86 89 113 124 137
## 2011 2012
## 101 110
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.001
```

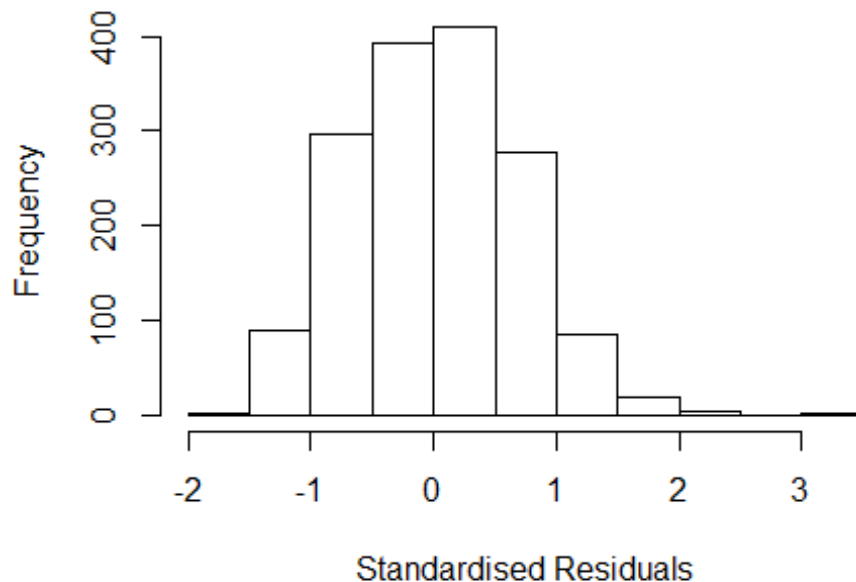


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.34, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 2.39719929057295"
## [1] "Male first author team size 2018 geometric mean: 2.7903550671996"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.2761858530381"
## [1] "Male last author team size 2018 geometric mean: 2.80299095387817"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.273 1 1.128
## LastAuthorFemale 1.268 1 1.126
## UniqueAuthors 1.201 4 1.023
## Year 1.245 16 1.007
```

## Residuals from first and last author and team size



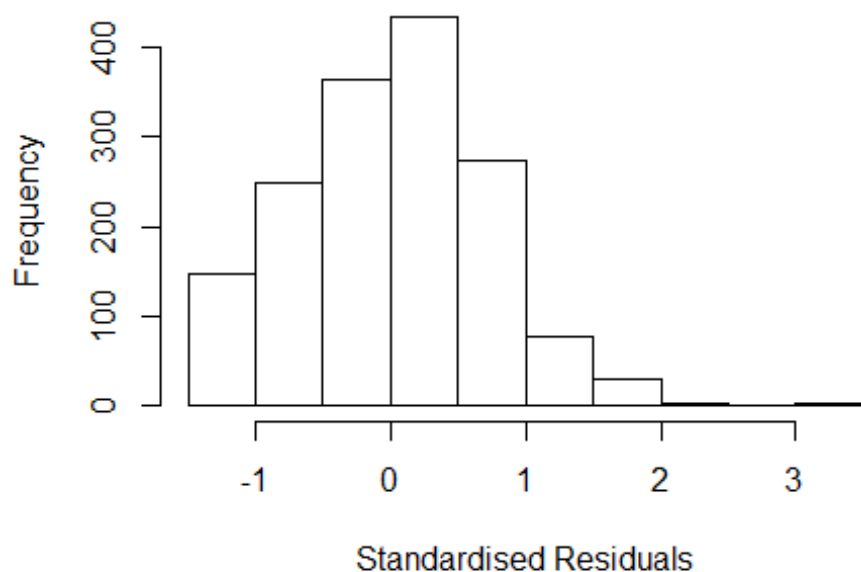
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3548 79954549466 4.139 2009      1204      2      3.402
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5460 -0.4790  0.0113  0.4866  3.4016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0947    0.0630   17.37 < 2e-16 ***
## FirstAuthorFemale1 -0.0801    0.0494   -1.62   0.105
## LastAuthorFemale1 -0.0446    0.0504   -0.88   0.377
## UniqueAuthors2     0.2722    0.0427    6.38 2.3e-10 ***
## UniqueAuthors3     0.3942    0.0575    6.86 9.8e-12 ***
## UniqueAuthors4     0.3717    0.0697    5.33 1.1e-07 ***
## UniqueAuthors5     0.6922    0.0699    9.90 < 2e-16 ***
## Year1997           0.0471    0.0914    0.52   0.606
## Year1998           0.0133    0.0995    0.13   0.894
## Year1999           0.0546    0.0896    0.61   0.542
```

```

## Year2000          0.0904      0.0893      1.01      0.312
## Year2001          -0.1368     0.1144     -1.20     0.232
## Year2002          -0.0724     0.0941     -0.77     0.442
## Year2003          -0.0370     0.0853     -0.43     0.665
## Year2004          -0.1433     0.0950     -1.51     0.132
## Year2005          -0.1562     0.0970     -1.61     0.107
## Year2006          -0.1262     0.1010     -1.25     0.212
## Year2007          -0.2409     0.1054     -2.29     0.022 *
## Year2008          -0.2382     0.0945     -2.52     0.012 *
## Year2009          -0.3573     0.0890     -4.02     6.2e-05 ***
## Year2010          -0.2606     0.0904     -2.88     0.004 **
## Year2011          -0.1744     0.0970     -1.80     0.073 .
## Year2012          -0.0826     0.0939     -0.88     0.379
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.0971
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1118 is an outlier with |weight| = 0 ( < 6.3e-05);
## 135 weights are ~= 1. The remaining 1443 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.199  0.879  0.949   0.918  0.985   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          6.33e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500          50          2          1          1000          200
## trace.lev mts compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.255 1          1.120
## LastAuthorFemale  1.250 1          1.118
## Year              1.057 16          1.002

```

## Residuals from first and last author



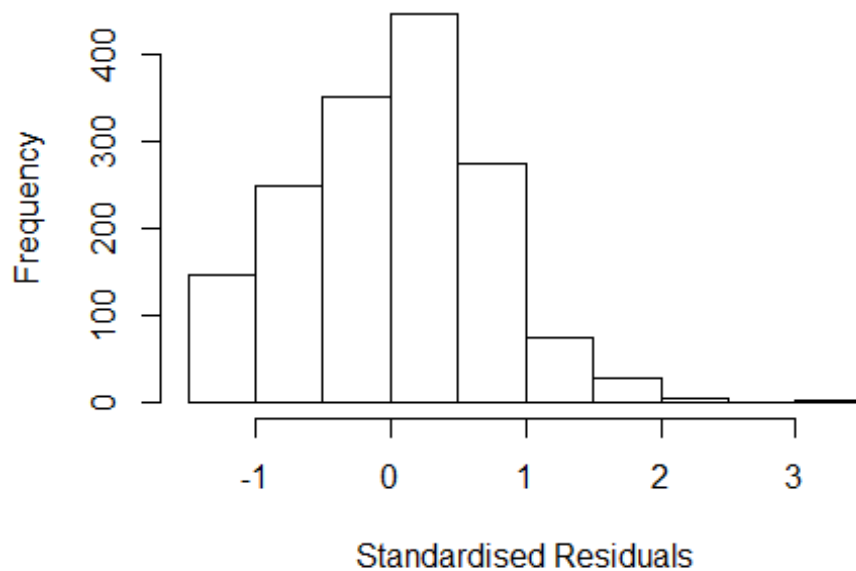
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3548 79954549466 4.139 2009      1204      2      3.271
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3221 -0.4988  0.0338  0.4822  3.2706
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21573    0.06030   20.16 < 2e-16 ***
## FirstAuthorFemale1 -0.03850    0.05080   -0.76  0.44864
## LastAuthorFemale1 -0.09203    0.05172   -1.78  0.07537 .
## Year1997         0.08316    0.09213    0.90  0.36683
## Year1998        -0.00398    0.10213   -0.04  0.96891
## Year1999         0.09365    0.09208    1.02  0.30928
## Year2000         0.10632    0.09040    1.18  0.23972
## Year2001        -0.05310    0.10988   -0.48  0.62898
## Year2002        -0.01349    0.09383   -0.14  0.88570
## Year2003         0.02870    0.08627    0.33  0.73946
## Year2004        -0.05851    0.09549   -0.61  0.54012
## Year2005        -0.08847    0.10035   -0.88  0.37811
```

```

## Year2006          -0.02012      0.09723    -0.21  0.83608
## Year2007          -0.15823      0.10677    -1.48  0.13857
## Year2008          -0.19945      0.10152    -1.96  0.04964 *
## Year2009          -0.34730      0.09158    -3.79  0.00016 ***
## Year2010          -0.22028      0.09502    -2.32  0.02057 *
## Year2011          -0.07161      0.10287    -0.70  0.48644
## Year2012          -0.01551      0.09818    -0.16  0.87452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.724
## Multiple R-squared:  0.0328, Adjusted R-squared:  0.0216
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0049 0.8710 0.9500 0.9120 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1      1.016
## Year              1.032 16      1.001

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3548 79954549466 4.139 2009      1204      2      3.271
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3109 -0.5019  0.0412  0.4850  3.2753
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20869    0.05978   20.22 < 2e-16 ***
## FirstAuthorFemale1 -0.08473    0.04664   -1.82  0.06946 .
## Year1997         0.08112    0.09178    0.88  0.37689
## Year1998        -0.00229    0.10196   -0.02  0.98209
## Year1999         0.09451    0.09200    1.03  0.30443
## Year2000         0.10218    0.09060    1.13  0.25958
## Year2001        -0.05026    0.10931   -0.46  0.64577
## Year2002        -0.00875    0.09371   -0.09  0.92565
## Year2003         0.02587    0.08554    0.30  0.76236
## Year2004        -0.05565    0.09523   -0.58  0.55908
## Year2005        -0.09578    0.10012   -0.96  0.33889
## Year2006        -0.01787    0.09730   -0.18  0.85428
```

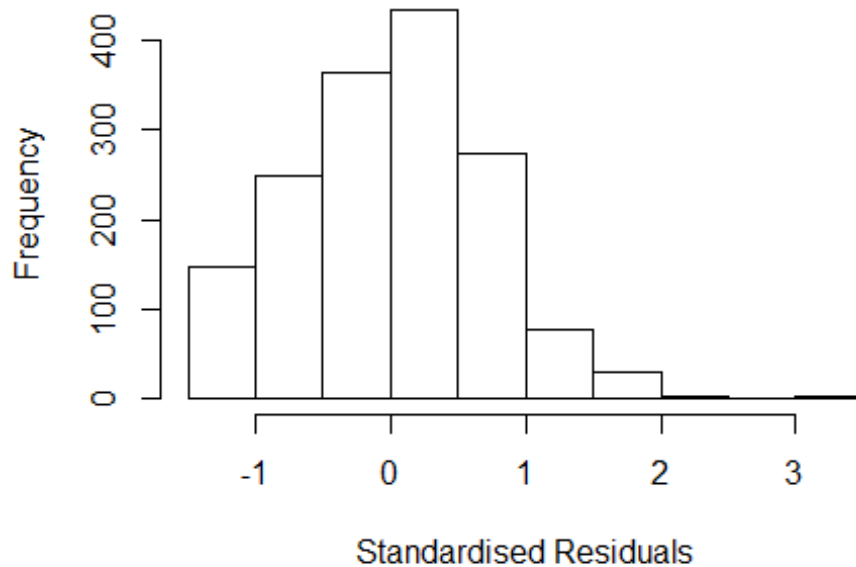


```

## Year2007          -0.16286    0.10682   -1.52  0.12754
## Year2008          -0.19833    0.10147   -1.95  0.05081 .
## Year2009          -0.34499    0.09158   -3.77  0.00017 ***
## Year2010          -0.22033    0.09496   -2.32  0.02046 *
## Year2011          -0.07395    0.10312   -0.72  0.47343
## Year2012          -0.02094    0.09792   -0.21  0.83067
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.725
## Multiple R-squared:  0.031, Adjusted R-squared:  0.0204
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 140 weights are ~= 1. The remaining 1439 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0049 0.8690 0.9500 0.9130 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1          1.014
## Year            1.028 16          1.001

```

## Residuals from last author



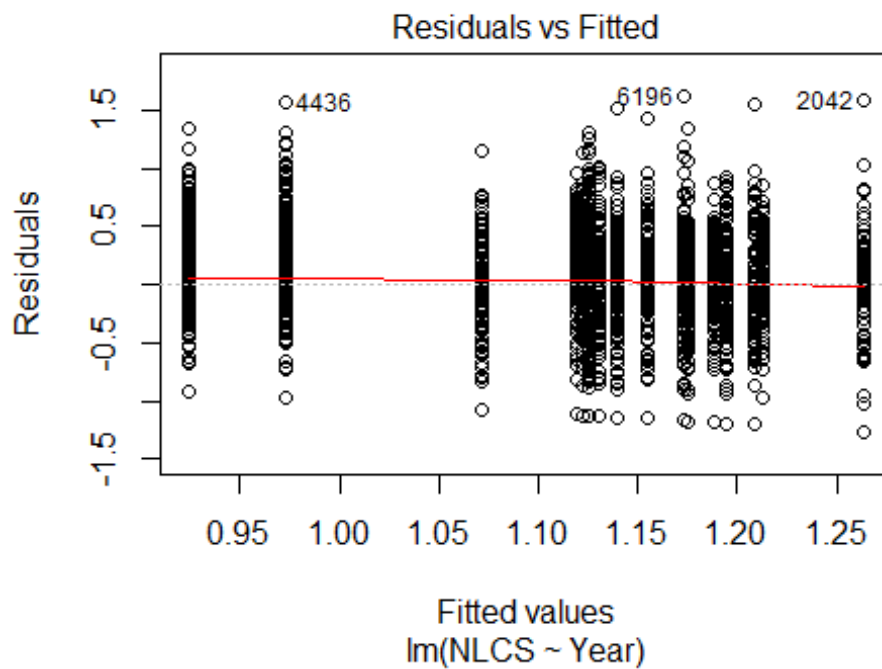
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3548 79954549466 4.139 2009      1204      2      3.271
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3201 -0.5042  0.0319  0.4869  3.2760
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21505    0.06037   20.13 < 2e-16 ***
## LastAuthorFemale1 -0.11274    0.04738   -2.38  0.01746 *
## Year1997         0.08007    0.09215    0.87  0.38499
## Year1998        -0.00768    0.10200   -0.08  0.93997
## Year1999         0.08997    0.09211    0.98  0.32886
## Year2000         0.10509    0.09024    1.16  0.24434
## Year2001        -0.05890    0.11022   -0.53  0.59315
## Year2002        -0.01779    0.09385   -0.19  0.84969
## Year2003         0.02435    0.08602    0.28  0.77719
## Year2004        -0.06316    0.09526   -0.66  0.50736
## Year2005        -0.09098    0.10008   -0.91  0.36348
## Year2006        -0.02444    0.09715   -0.25  0.80144
```

```

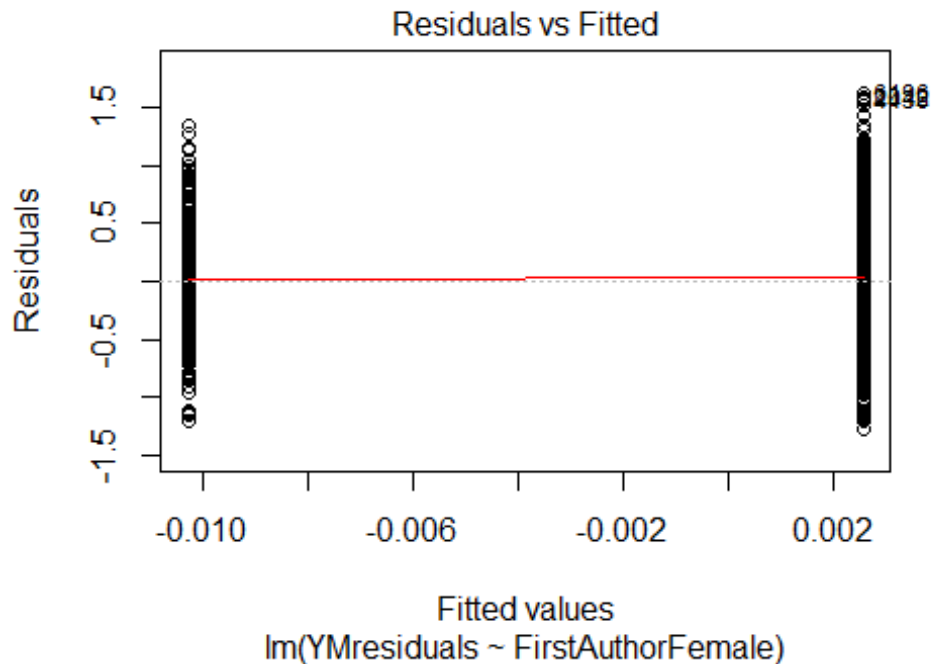
## Year2007          -0.16189      0.10689    -1.51  0.13009
## Year2008          -0.20121      0.10149    -1.98  0.04759 *
## Year2009          -0.35202      0.09136    -3.85  0.00012 ***
## Year2010          -0.22217      0.09504    -2.34  0.01953 *
## Year2011          -0.07483      0.10286    -0.73  0.46702
## Year2012          -0.02025      0.09806    -0.21  0.83645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.725
## Multiple R-squared:  0.0323, Adjusted R-squared:  0.0218
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1440 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.005  0.870  0.951  0.913  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.33e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1579"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1901"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 293 336 279 289 288 320 321 308 303 266 325 347 282 248 284
## 2011 2012
## 254 259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 111 124 122 104 82 132 130 142 153 186 205 156 124 162
## 2011 2012

```

```
## 121 138
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 87 100 113 114 97 75 117 121 134 122 169 177 136 109 152
## 2011 2012
## 112 128
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 84, df = 16, p-value = 3e-11
```

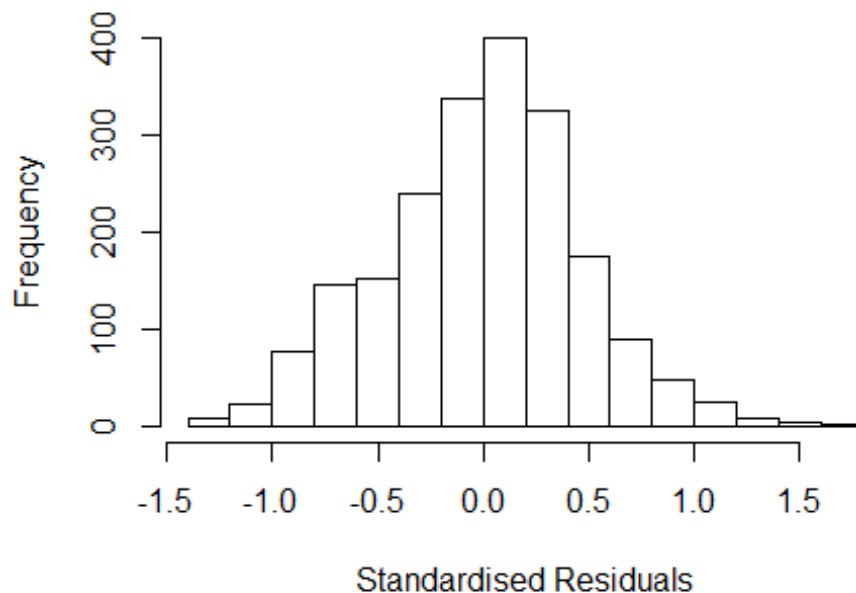


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 2.73344741681578"
## [1] "Male first author team size 2018 geometric mean: 2.805073045383"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 690, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.62862448279474"
## [1] "Male last author team size 2018 geometric mean: 2.83359336773911"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.072 1          1.035
## LastAuthorFemale  1.084 1          1.041
## UniqueAuthors    1.312 4          1.035
## Year              1.363 16         1.010
```

## Residuals from first and last author and team size



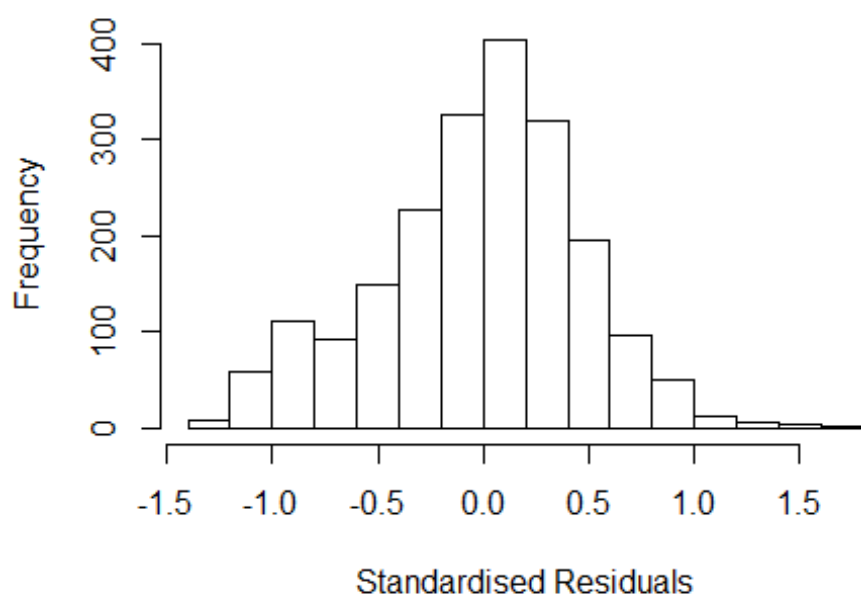
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2722 -0.2895  0.0294  0.2987  1.7966
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04407    0.04951   21.09 < 2e-16 ***
## FirstAuthorFemale1 -0.03687    0.02614   -1.41  0.1586
## LastAuthorFemale1 -0.00945    0.02976   -0.32  0.7508
## UniqueAuthors2     0.22814    0.03090    7.38 2.2e-13 ***
## UniqueAuthors3     0.26527    0.03237    8.20 4.4e-16 ***
## UniqueAuthors4     0.29428    0.03679    8.00 2.1e-15 ***
## UniqueAuthors5     0.34462    0.03518    9.79 < 2e-16 ***
## Year1997         -0.01620    0.06640   -0.24  0.8072
## Year1998         -0.11595    0.06579   -1.76  0.0782 .
## Year1999         -0.05486    0.06278   -0.87  0.3823
```

```

## Year2000      -0.02235      0.06005      -0.37      0.7098
## Year2001      -0.01163      0.06885      -0.17      0.8659
## Year2002      -0.01778      0.06380      -0.28      0.7805
## Year2003      -0.13555      0.06263      -2.16      0.0305 *
## Year2004      -0.09255      0.05859      -1.58      0.1144
## Year2005      -0.09368      0.06992      -1.34      0.1805
## Year2006      -0.24999      0.07171      -3.49      0.0005 ***
## Year2007      -0.28706      0.06747      -4.25      2.2e-05 ***
## Year2008      -0.10006      0.06694      -1.49      0.1352
## Year2009      -0.00495      0.06052      -0.08      0.9348
## Year2010      -0.04768      0.05850      -0.81      0.4152
## Year2011      -0.08820      0.06561      -1.34      0.1790
## Year2012      -0.08083      0.06195      -1.30      0.1921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.106, Adjusted R-squared:  0.0963
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1895 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0457 0.8520 0.9490 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.85e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.055 1      1.027
## LastAuthorFemale  1.071 1      1.035
## Year              1.106 16      1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2419 -0.3052  0.0295  0.3005  1.6097
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18029    0.05035   23.44 < 2e-16 ***
## FirstAuthorFemale1 -0.01113    0.02648   -0.42  0.67431
## LastAuthorFemale1 -0.00312    0.03092   -0.10  0.91975
## Year1997        -0.00811    0.06821   -0.12  0.90538
## Year1998        -0.06146    0.06854   -0.90  0.36999
## Year1999        -0.04070    0.06588   -0.62  0.53683
## Year2000         0.03815    0.06313    0.60  0.54575
## Year2001         0.06157    0.06925    0.89  0.37408
## Year2002         0.03206    0.06533    0.49  0.62368
## Year2003        -0.09324    0.06436   -1.45  0.14759
## Year2004        -0.04616    0.06195   -0.75  0.45632
## Year2005        -0.03109    0.07529   -0.41  0.67971
```

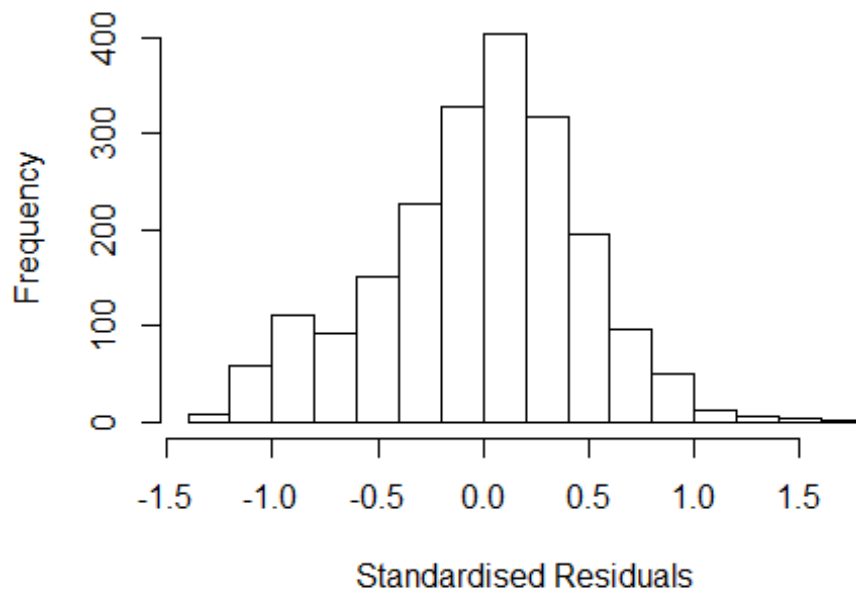


```

## Year2006          -0.21938      0.07821    -2.81  0.00508 **
## Year2007          -0.25586      0.07364    -3.47  0.00052 ***
## Year2008          -0.02676      0.06924    -0.39  0.69921
## Year2009           0.04787      0.06290     0.76  0.44680
## Year2010           0.00300      0.06033     0.05  0.96033
## Year2011          -0.02100      0.06671    -0.31  0.75302
## Year2012          -0.00236      0.06284    -0.04  0.97003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0271
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1893 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.149  0.850  0.948  0.888  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.049 1      1.024
## Year              1.049 16      1.001

```

## Residuals from first author



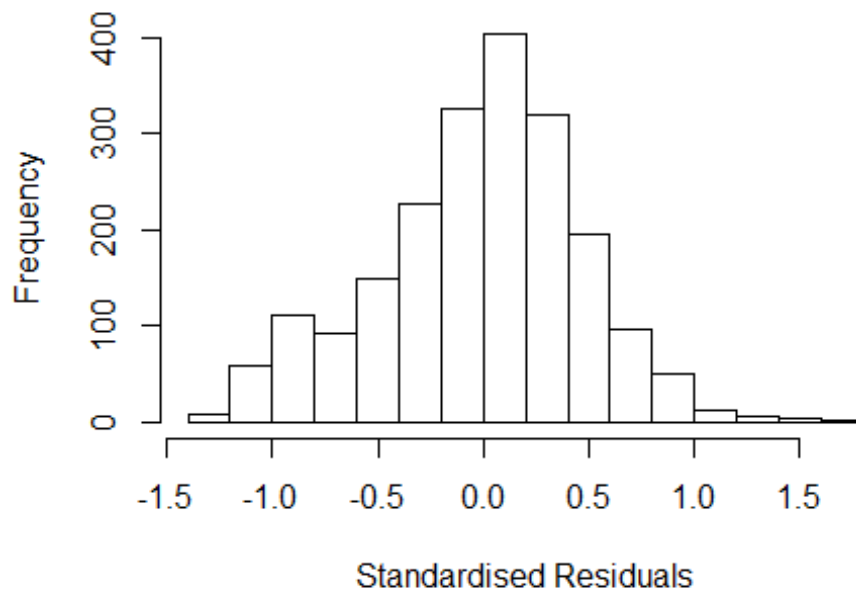
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2418 -0.3049 0.0298 0.3002 1.6103
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17997 0.05020 23.51 < 2e-16 ***
## FirstAuthorFemale1 -0.01182 0.02670 -0.44 0.65800
## Year1997 -0.00819 0.06820 -0.12 0.90442
## Year1998 -0.06132 0.06847 -0.90 0.37060
## Year1999 -0.04055 0.06582 -0.62 0.53788
## Year2000 0.03792 0.06304 0.60 0.54760
## Year2001 0.06185 0.06912 0.89 0.37101
## Year2002 0.03208 0.06530 0.49 0.62327
## Year2003 -0.09319 0.06437 -1.45 0.14784
## Year2004 -0.04621 0.06192 -0.75 0.45564
## Year2005 -0.03083 0.07512 -0.41 0.68159
## Year2006 -0.21946 0.07821 -2.81 0.00506 **
```

```

## Year2007          -0.25599    0.07364   -3.48  0.00052 ***
## Year2008          -0.02680    0.06924   -0.39  0.69874
## Year2009           0.04778    0.06285    0.76  0.44727
## Year2010           0.00276    0.06019    0.05  0.96345
## Year2011          -0.02096    0.06670   -0.31  0.75334
## Year2012          -0.00232    0.06285   -0.04  0.97053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0276
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1895 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.149  0.851  0.948  0.888  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.06 1          1.030
## Year              1.06 16          1.002

```

## Residuals from last author



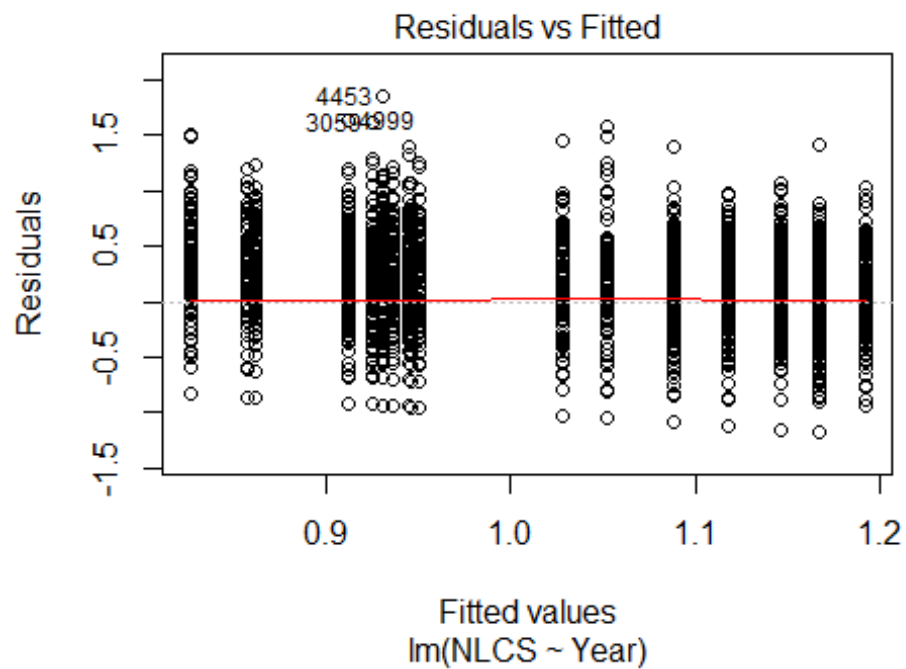
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2404 -0.3034 0.0306 0.3020 1.6115
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17963 0.05029 23.46 <2e-16 ***
## LastAuthorFemale1 -0.00627 0.03115 -0.20 0.8406
## Year1997 -0.00845 0.06818 -0.12 0.9013
## Year1998 -0.06296 0.06844 -0.92 0.3577
## Year1999 -0.04110 0.06590 -0.62 0.5329
## Year2000 0.03738 0.06316 0.59 0.5540
## Year2001 0.06076 0.06922 0.88 0.3802
## Year2002 0.03046 0.06509 0.47 0.6398
## Year2003 -0.09399 0.06440 -1.46 0.1446
## Year2004 -0.04723 0.06190 -0.76 0.4455
## Year2005 -0.03258 0.07529 -0.43 0.6653
## Year2006 -0.22049 0.07826 -2.82 0.0049 **
```

```

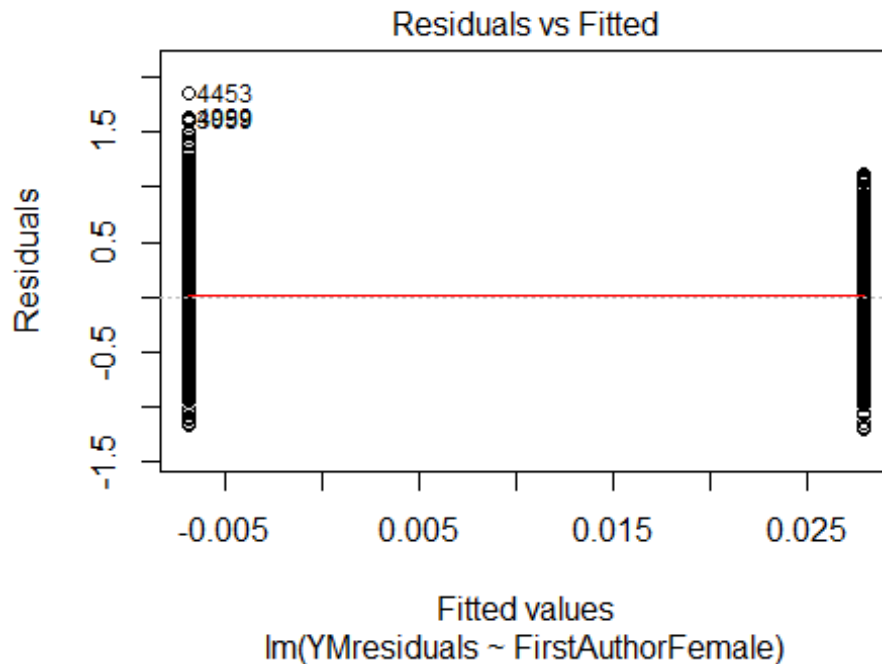
## Year2007          -0.25677      0.07361    -3.49    0.0005 ***
## Year2008          -0.02817      0.06918    -0.41    0.6839
## Year2009           0.04688      0.06294     0.74    0.4564
## Year2010           0.00187      0.06029     0.03    0.9752
## Year2011          -0.02265      0.06664    -0.34    0.7340
## Year2012          -0.00496      0.06249    -0.08    0.9368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.0355, Adjusted R-squared:  0.0275
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 171 weights are ~= 1. The remaining 1892 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.149  0.850  0.948  0.888  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2063"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1902"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 292 339 332 347 408 425 382 384 386 382 365 367 336 364 359
## 2011 2012
## 364 379
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 119 118 148 141 112 170 190 197 223 223 183 202 192 189
## 2011 2012

```

```
## 166 205
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 99 109 113 136 125 105 155 180 174 178 188 142 169 171 173
## 2011 2012
## 146 185
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 80, df = 16, p-value = 2e-10
```

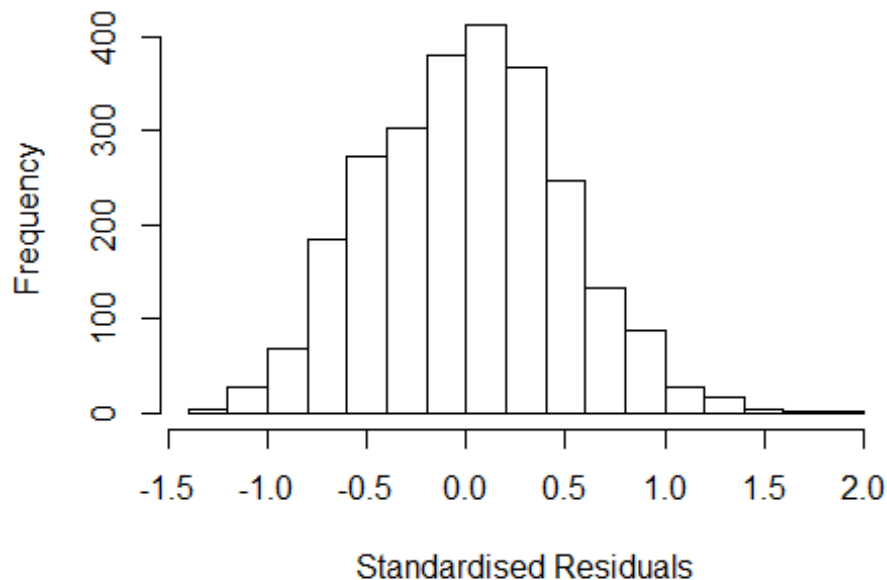


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 3e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.0909898105372"
## [1] "Male first author team size 2018 geometric mean: 3.09679927971817"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5100, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.82041127002095"
## [1] "Male last author team size 2018 geometric mean: 3.15400474510308"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.085 1          1.042
## LastAuthorFemale  1.076 1          1.037
## UniqueAuthors    1.191 4          1.022
## Year             1.224 16          1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3082 -0.3502 0.0201 0.3265 1.8029
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.68140 0.05236 13.01 <2e-16 ***
## FirstAuthorFemale1 -0.00550 0.02479 -0.22 0.824
## LastAuthorFemale1 0.00766 0.02959 0.26 0.796
## UniqueAuthors2 0.41146 0.02848 14.45 <2e-16 ***
## UniqueAuthors3 0.46624 0.03007 15.51 <2e-16 ***
## UniqueAuthors4 0.52266 0.03471 15.06 <2e-16 ***
## UniqueAuthors5 0.60292 0.03239 18.62 <2e-16 ***
## Year1997 0.17461 0.07666 2.28 0.023 *
## Year1998 -0.00181 0.07039 -0.03 0.979
## Year1999 0.08574 0.06719 1.28 0.202
```

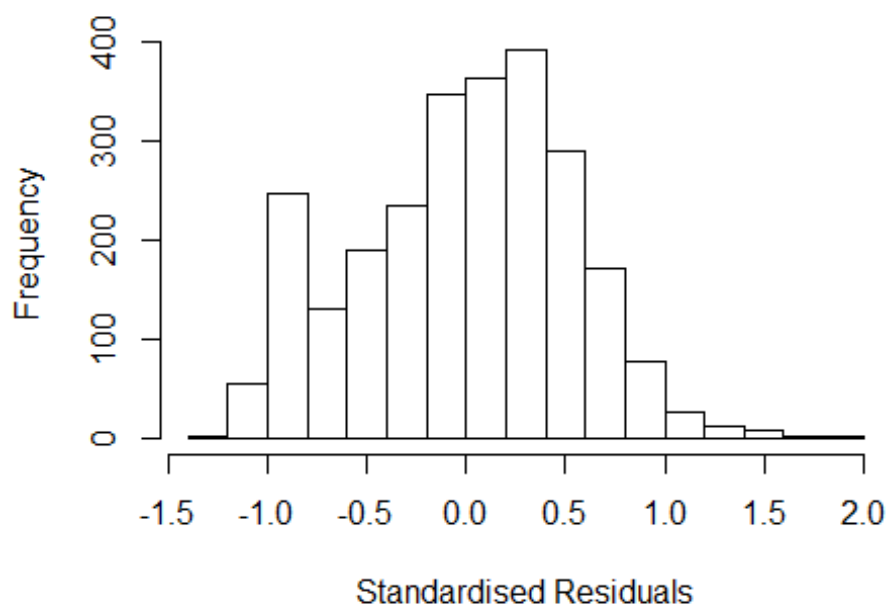


```

## Year2000      -0.02536    0.06673   -0.38    0.704
## Year2001      -0.10818    0.07405   -1.46    0.144
## Year2002      -0.07986    0.06644   -1.20    0.230
## Year2003      -0.15334    0.06635   -2.31    0.021 *
## Year2004      -0.12965    0.06287   -2.06    0.039 *
## Year2005      -0.07239    0.06426   -1.13    0.260
## Year2006      -0.09744    0.06237   -1.56    0.118
## Year2007      -0.10382    0.06526   -1.59    0.112
## Year2008       0.09800    0.06233    1.57    0.116
## Year2009       0.07688    0.06342    1.21    0.225
## Year2010       0.10963    0.06096    1.80    0.072 .
## Year2011       0.12544    0.06035    2.08    0.038 *
## Year2012       0.12260    0.06202    1.98    0.048 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.223, Adjusted R-squared:  0.216
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 189 weights are ~= 1. The remaining 2359 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.143  0.874  0.949  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.092 1      1.045
## LastAuthorFemale  1.090 1      1.044
## Year              1.063 16      1.002

```

## Residuals from first and last author

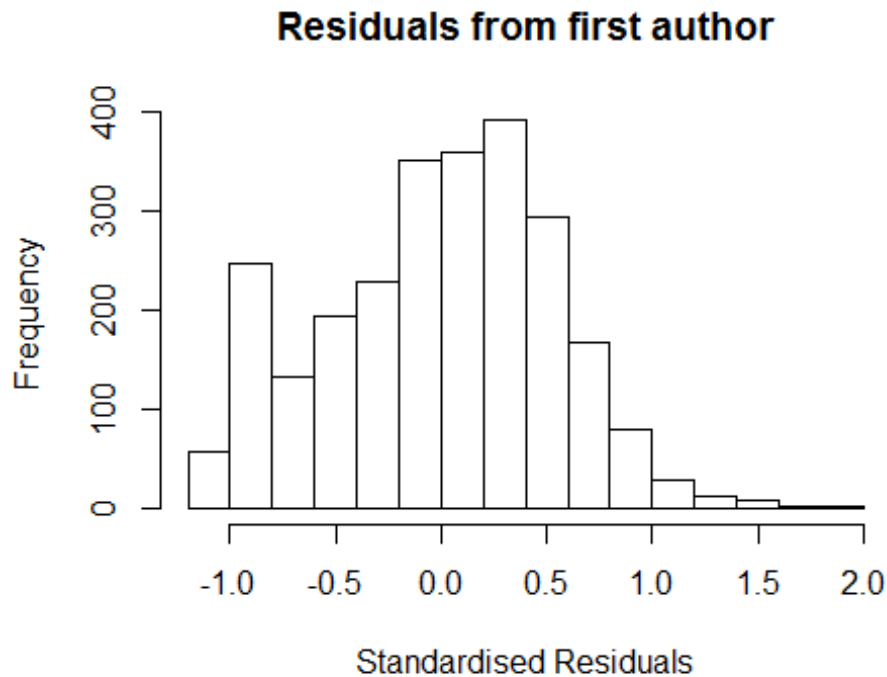


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2033 -0.3866 0.0378 0.3732 1.9011
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90362 0.06167 14.65 < 2e-16 ***
## FirstAuthorFemale1 0.06489 0.02651 2.45 0.01443 *
## LastAuthorFemale1 -0.02157 0.03153 -0.68 0.49406
## Year1997 0.15019 0.08845 1.70 0.08960 .
## Year1998 0.03498 0.08251 0.42 0.67163
## Year1999 0.09788 0.07929 1.23 0.21714
## Year2000 -0.00930 0.08375 -0.11 0.91163
## Year2001 -0.06584 0.09257 -0.71 0.47696
## Year2002 -0.00551 0.08085 -0.07 0.94571
## Year2003 -0.12053 0.08111 -1.49 0.13739
## Year2004 -0.08971 0.07869 -1.14 0.25438
## Year2005 -0.02774 0.07811 -0.36 0.72246
```

```

## Year2006      -0.02651    0.07538   -0.35  0.72512
## Year2007      -0.00588    0.07942   -0.07  0.94095
## Year2008       0.20205    0.07302    2.77  0.00570 **
## Year2009       0.18241    0.07107    2.57  0.01032 *
## Year2010       0.23476    0.06980    3.36  0.00078 ***
## Year2011       0.25262    0.06904    3.66  0.00026 ***
## Year2012       0.23344    0.07019    3.33  0.00089 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0537, Adjusted R-squared:  0.047
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 215 weights are ~= 1. The remaining 2333 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.179  0.860  0.946  0.907  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1      1.018
## Year              1.036 16      1.001

```

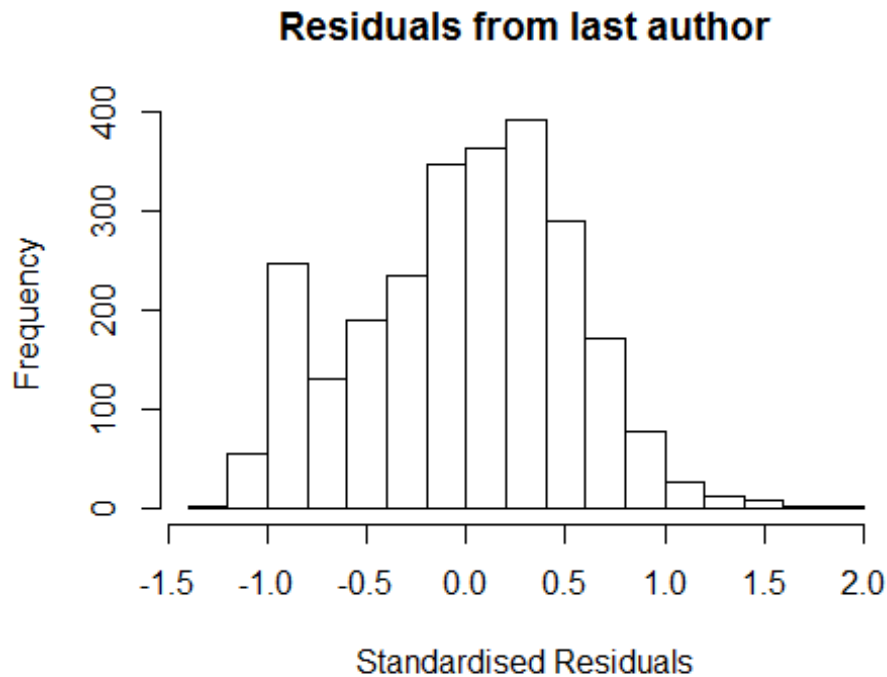


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1947 -0.3871 0.0391 0.3725 1.9028
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90251 0.06178 14.61 < 2e-16 ***
## FirstAuthorFemale1 0.05897 0.02608 2.26 0.02382 *
## Year1997 0.14977 0.08865 1.69 0.09126 .
## Year1998 0.03516 0.08267 0.43 0.67069
## Year1999 0.09757 0.07947 1.23 0.21965
## Year2000 -0.01042 0.08385 -0.12 0.90114
## Year2001 -0.06534 0.09288 -0.70 0.48180
## Year2002 -0.00608 0.08112 -0.07 0.94029
## Year2003 -0.12133 0.08111 -1.50 0.13483
## Year2004 -0.08932 0.07888 -1.13 0.25763
## Year2005 -0.02833 0.07821 -0.36 0.71724
## Year2006 -0.02672 0.07550 -0.35 0.72342
```

```

## Year2007          -0.00643    0.07963   -0.08  0.93563
## Year2008          0.20125    0.07311    2.75  0.00595 **
## Year2009          0.18159    0.07114    2.55  0.01075 *
## Year2010          0.23317    0.06975    3.34  0.00084 ***
## Year2011          0.25105    0.06901    3.64  0.00028 ***
## Year2012          0.23202    0.07022    3.30  0.00097 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.0537, Adjusted R-squared:  0.0474
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 223 weights are ~= 1. The remaining 2325 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.172  0.857  0.945  0.905  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1          1.014
## Year            1.028 16          1.001

```



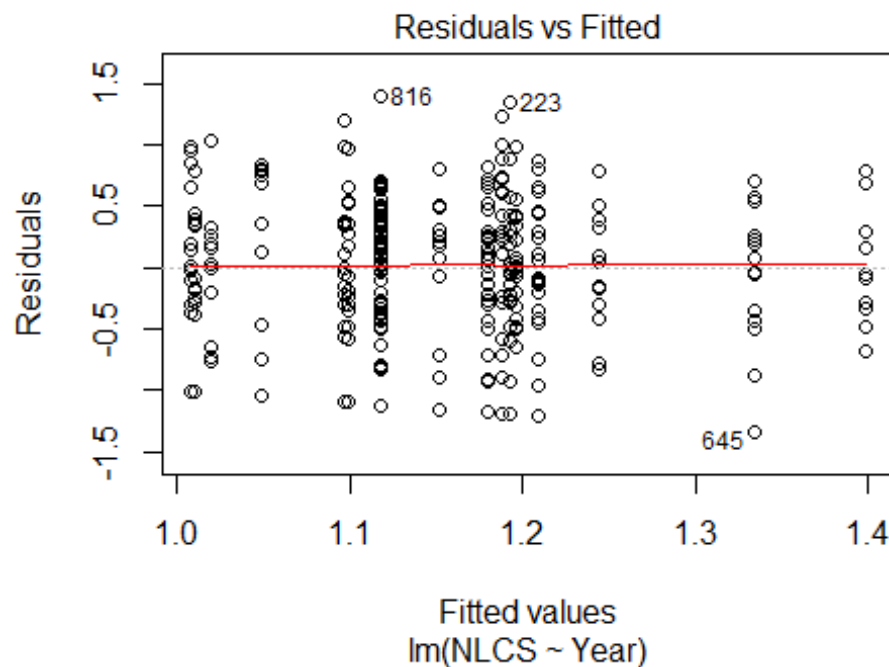
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1567 -0.3834 0.0376 0.3699 1.8918
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.909495 0.061826 14.71 < 2e-16 ***
## LastAuthorFemale1 0.002402 0.030276 0.08 0.93677
## Year1997 0.148265 0.088472 1.68 0.09389 .
## Year1998 0.042300 0.082701 0.51 0.60906
## Year1999 0.101574 0.079697 1.27 0.20260
## Year2000 -0.005869 0.084015 -0.07 0.94432
## Year2001 -0.064321 0.093178 -0.69 0.49007
## Year2002 -0.000676 0.081043 -0.01 0.99335
## Year2003 -0.117732 0.081427 -1.45 0.14834
## Year2004 -0.089062 0.079195 -1.12 0.26087
## Year2005 -0.024316 0.078394 -0.31 0.75645
## Year2006 -0.023649 0.075685 -0.31 0.75472
```

```

## Year2007          -0.005104    0.079751    -0.06    0.94898
## Year2008          0.207120    0.073166     2.83    0.00468 **
## Year2009          0.192480    0.071074     2.71    0.00681 **
## Year2010          0.240759    0.069850     3.45    0.00058 ***
## Year2011          0.255571    0.068899     3.71    0.00021 ***
## Year2012          0.244843    0.070081     3.49    0.00048 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.53
## Multiple R-squared:  0.0518, Adjusted R-squared:  0.0454
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 210 weights are ~= 1. The remaining 2338 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.176  0.860  0.944  0.905  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2548"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1903"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   43   45   27   43   49   33   32   26   29   26   29   38   45   33
## 2011 2012
##   45   74
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10   19   23   12   14   22   24   15   12   14   14   12   23   31   18
## 2011 2012

```

```
## 27 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 18 22 11 14 21 21 15 10 12 14 10 22 28 17
## 2011 2012
## 26 45
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```

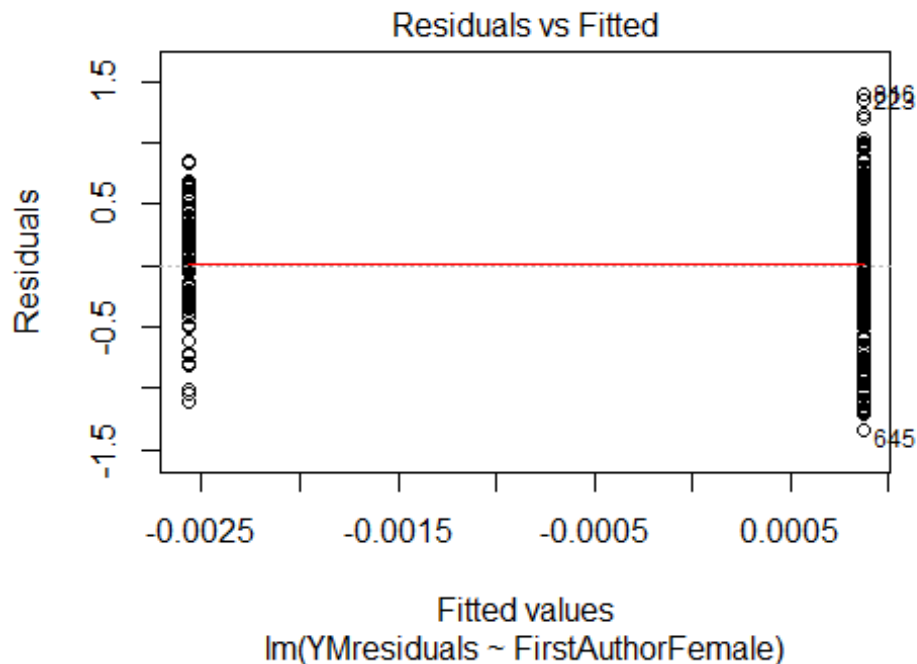


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.5, df = 1, p-value = 0.01
## [1] "Female first author team size 2018 geometric mean: 3.15689811951386"
## [1] "Male first author team size 2018 geometric mean: 2.33406195690484"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 160, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.71080601082953"
## [1] "Male last author team size 2018 geometric mean: 2.5906992575107"

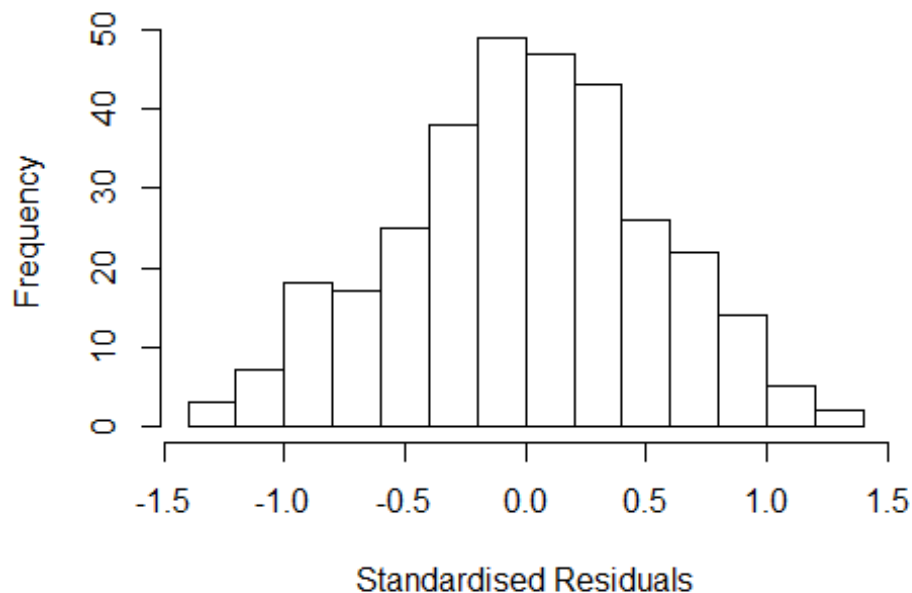
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 60, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.518	1	1.232
LastAuthorFemale	1.687	1	1.299
UniqueAuthors	2.681	4	1.131
Year	3.276	16	1.038

## Residuals from first and last author and team size



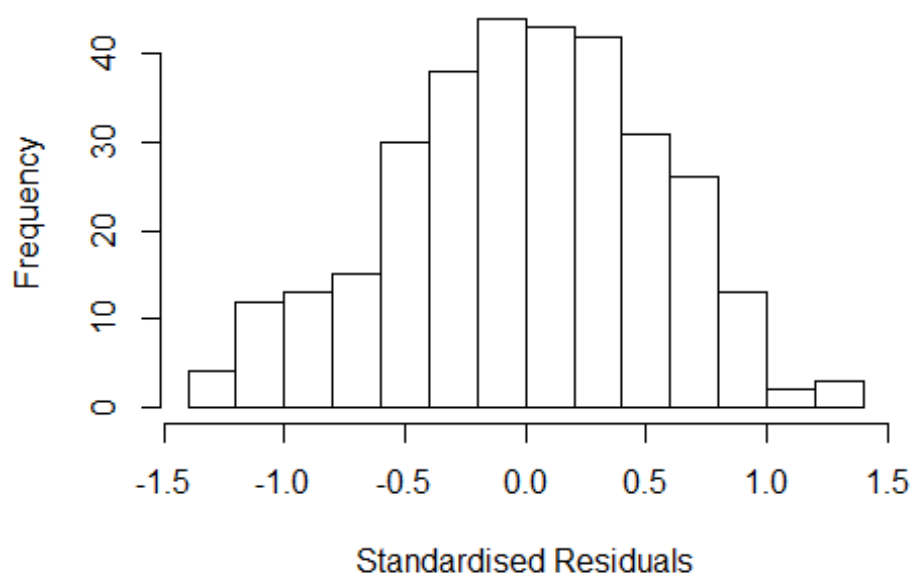
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27295 -0.34128  0.00462  0.36143  1.30825
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3060    0.1690   7.73 1.8e-13 ***
## FirstAuthorFemale1 -0.0764    0.0766  -1.00  0.3195
## LastAuthorFemale1  0.0666    0.0893   0.75  0.4559
## UniqueAuthors2     0.1000    0.0892   1.12  0.2634
## UniqueAuthors3     0.3090    0.1026   3.01  0.0028 **
## UniqueAuthors4     0.2046    0.1077   1.90  0.0583 .
## UniqueAuthors5     0.4997    0.1111   4.50 1.0e-05 ***
## Year1997          -0.3478    0.2024  -1.72  0.0868 .
## Year1998          -0.1943    0.2492  -0.78  0.4361
## Year1999          -0.4214    0.3757  -1.12  0.2629
```

```

## Year2000          -0.1666      0.2139   -0.78    0.4367
## Year2001          -0.2287      0.2036   -1.12    0.2623
## Year2002          -0.1832      0.2019   -0.91    0.3648
## Year2003          -0.3342      0.2206   -1.52    0.1308
## Year2004          -0.3578      0.2313   -1.55    0.1229
## Year2005          -0.4501      0.2990   -1.51    0.1334
## Year2006          -0.4363      0.1985   -2.20    0.0287 *
## Year2007          -0.3443      0.2571   -1.34    0.1814
## Year2008          -0.1453      0.1832   -0.79    0.4283
## Year2009          -0.1299      0.1881   -0.69    0.4905
## Year2010          -0.0331      0.2084   -0.16    0.8740
## Year2011          -0.3469      0.1838   -1.89    0.0601 .
## Year2012          -0.3552      0.1864   -1.91    0.0577 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0894, Adjusted R-squared:  0.0211
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 289 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.539  0.872  0.954   0.911  0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.477 1      1.215
## LastAuthorFemale  1.453 1      1.205
## Year              1.279 16      1.008

```

## Residuals from first and last author



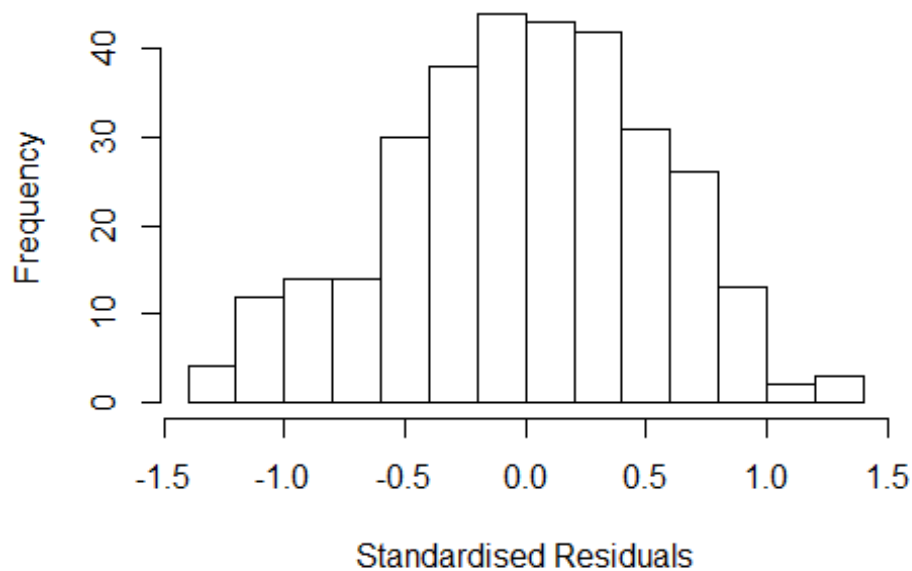
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3937 -0.3548 0.0101 0.3760 1.3270
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.39131 0.15609 8.91 <2e-16 ***
## FirstAuthorFemale1 -0.02299 0.07702 -0.30 0.766
## LastAuthorFemale1 -0.00772 0.08584 -0.09 0.928
## Year1997 -0.31519 0.20208 -1.56 0.120
## Year1998 -0.22830 0.24623 -0.93 0.355
## Year1999 -0.36684 0.39204 -0.94 0.350
## Year2000 -0.13219 0.20061 -0.66 0.510
## Year2001 -0.18529 0.19447 -0.95 0.341
## Year2002 -0.16451 0.19552 -0.84 0.401
## Year2003 -0.31264 0.21754 -1.44 0.152
## Year2004 -0.32955 0.22400 -1.47 0.142
## Year2005 -0.39140 0.31583 -1.24 0.216
```

```

## Year2006      -0.36965    0.19688   -1.88    0.061 .
## Year2007      -0.21950    0.27615   -0.79    0.427
## Year2008      -0.18820    0.17553   -1.07    0.285
## Year2009      -0.14298    0.17740   -0.81    0.421
## Year2010       0.00239    0.20270    0.01    0.991
## Year2011      -0.28060    0.18349   -1.53    0.127
## Year2012      -0.28615    0.17910   -1.60    0.111
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.033, Adjusted R-squared:  -0.0256
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 288 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.510  0.865   0.950   0.910   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.146 1      1.070
## Year              1.146 16      1.004

```

## Residuals from first author



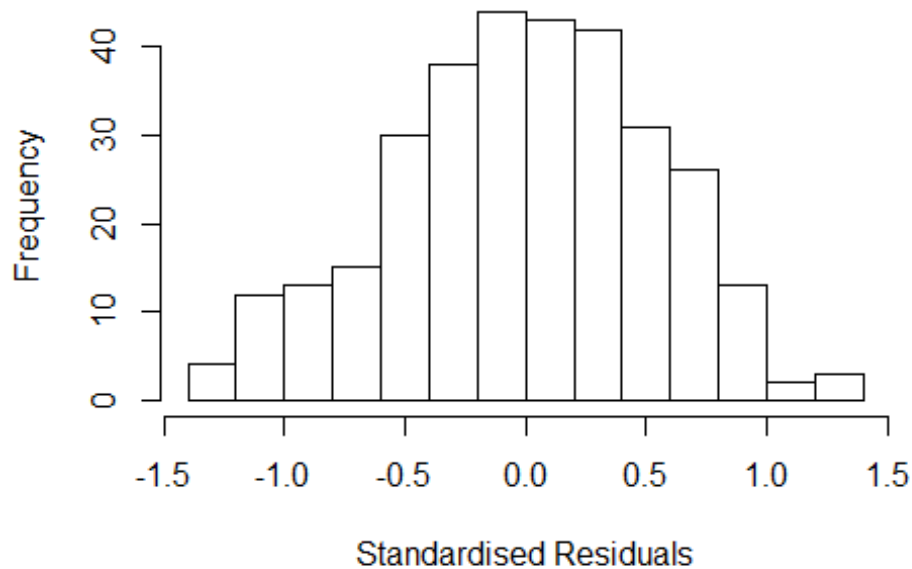
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3936 -0.3580 0.0107 0.3791 1.3276
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.38931 0.15481 8.97 <2e-16 ***
## FirstAuthorFemale1 -0.02661 0.06800 -0.39 0.696
## Year1997 -0.31330 0.20179 -1.55 0.122
## Year1998 -0.22583 0.24478 -0.92 0.357
## Year1999 -0.36640 0.39406 -0.93 0.353
## Year2000 -0.13051 0.20026 -0.65 0.515
## Year2001 -0.18394 0.19437 -0.95 0.345
## Year2002 -0.16348 0.19559 -0.84 0.404
## Year2003 -0.31044 0.21614 -1.44 0.152
## Year2004 -0.32735 0.22190 -1.48 0.141
## Year2005 -0.38954 0.31545 -1.23 0.218
## Year2006 -0.36800 0.19690 -1.87 0.063 .
```

```

## Year2007      -0.21973    0.27536   -0.80    0.426
## Year2008      -0.18754    0.17538   -1.07    0.286
## Year2009      -0.14179    0.17726   -0.80    0.424
## Year2010       0.00432    0.20287    0.02    0.983
## Year2011      -0.27947    0.18320   -1.53    0.128
## Year2012      -0.28482    0.17869   -1.59    0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.033, Adjusted R-squared:  -0.0221
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 288 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.509  0.865  0.950  0.910  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.12 1      1.058
## Year      1.12 16      1.004

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3862 -0.3532 0.0144 0.3799 1.3290
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.39261 0.15600 8.93 <2e-16 ***
## LastAuthorFemale1 -0.01988 0.07551 -0.26 0.793
## Year1997 -0.32090 0.20030 -1.60 0.110
## Year1998 -0.23260 0.24394 -0.95 0.341
## Year1999 -0.37060 0.38446 -0.96 0.336
## Year2000 -0.13727 0.19994 -0.69 0.493
## Year2001 -0.18856 0.19350 -0.97 0.331
## Year2002 -0.16743 0.19537 -0.86 0.392
## Year2003 -0.31851 0.21524 -1.48 0.140
## Year2004 -0.33424 0.22346 -1.50 0.136
## Year2005 -0.39314 0.31596 -1.24 0.214
## Year2006 -0.37174 0.19641 -1.89 0.059 .
```

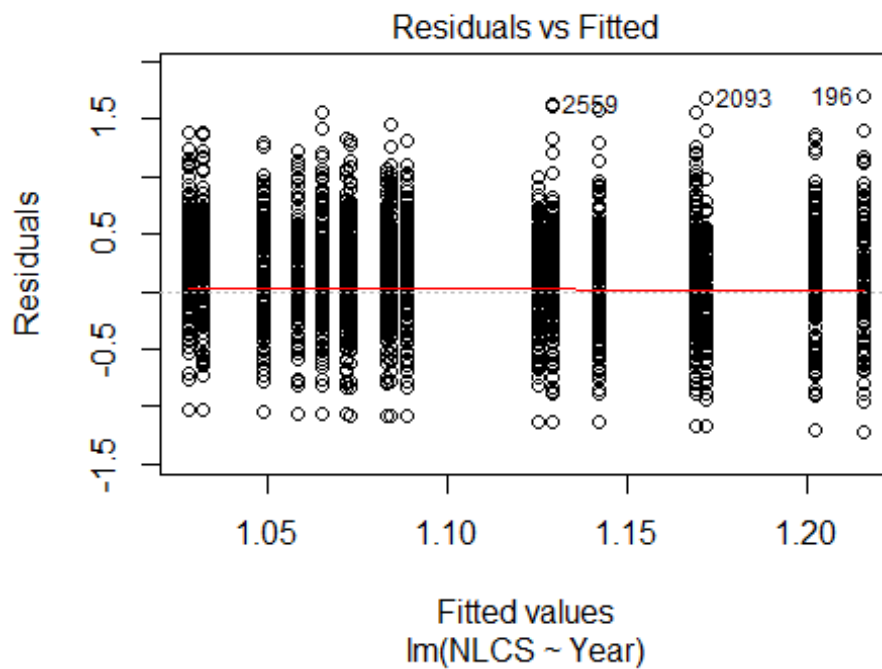


```

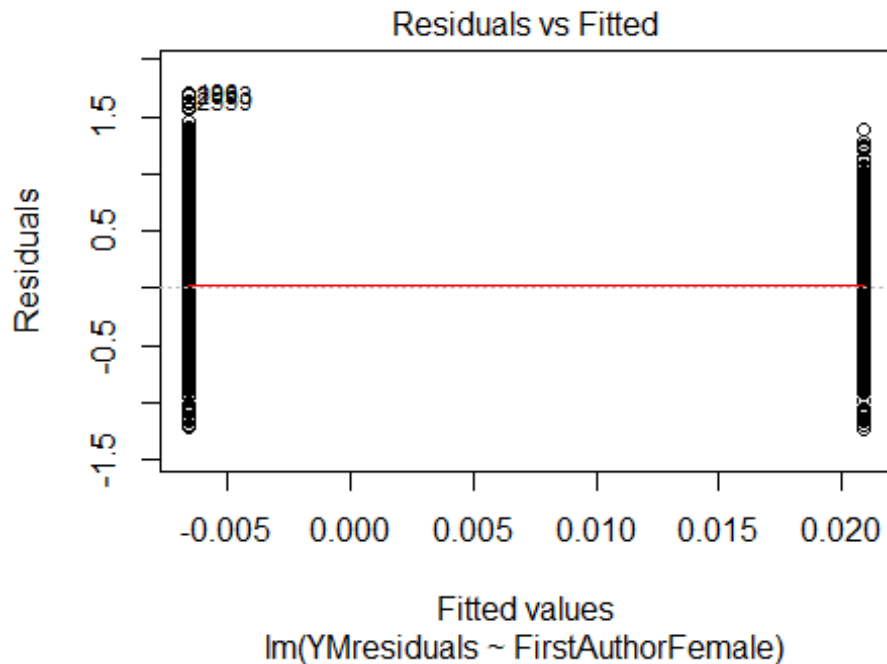
## Year2007          -0.22039      0.27668    -0.80      0.426
## Year2008          -0.19383      0.17436    -1.11      0.267
## Year2009          -0.14825      0.17670    -0.84      0.402
## Year2010          -0.00638      0.20020    -0.03      0.975
## Year2011          -0.28541      0.18417    -1.55      0.122
## Year2012          -0.29519      0.17682    -1.67      0.096 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.0326, Adjusted R-squared:  -0.0225
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 286 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.519  0.865   0.952   0.910   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.16e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 316"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1904"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  318  295  314  350  359  331  261  263  263  263  311  314  357  324  335
## 2011 2012
##  347  366
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  177  165  184  219  183  132  169  184  177  205  216  198  254  226  219
## 2011 2012

```

```
## 236 240
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 163 154 172 212 170 125 160 178 165 196 208 184 238 214 207
## 2011 2012
## 218 223
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.03
```

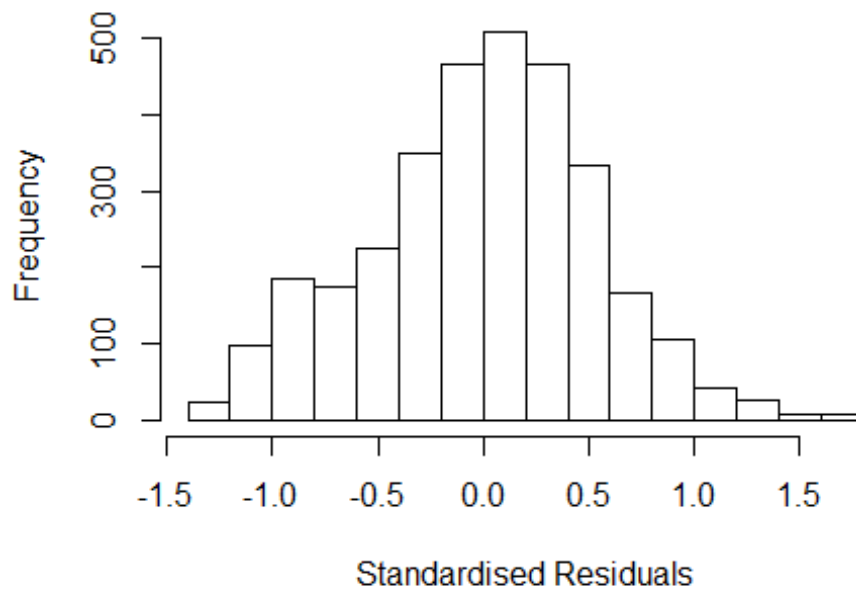


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.3, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.11998519938468"
## [1] "Male first author team size 2018 geometric mean: 2.07010714061886"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3300, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.68985620006978"
## [1] "Male last author team size 2018 geometric mean: 2.26495214728261"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.151 1      1.073
## LastAuthorFemale  1.143 1      1.069
## UniqueAuthors     1.174 4      1.020
## Year              1.225 16      1.006
```

## Residuals from first and last author and team size



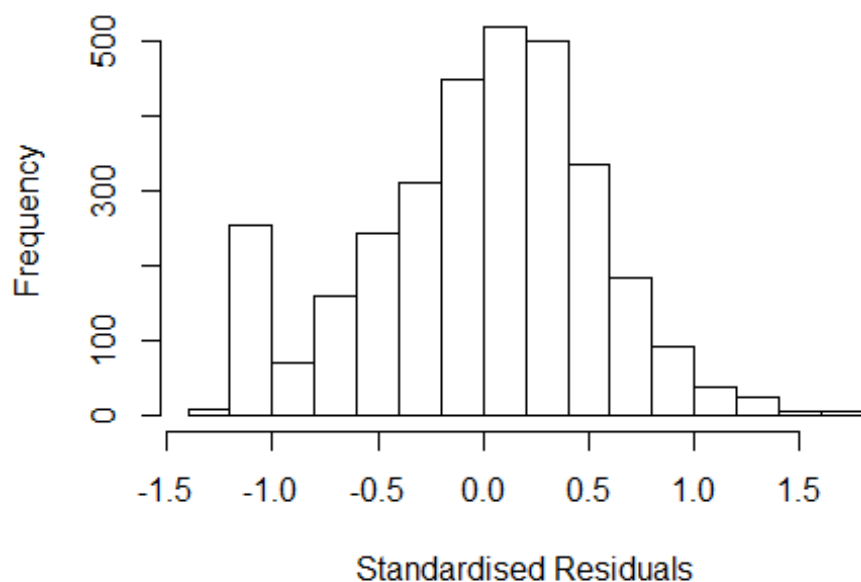
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3893 -0.3298 0.0266 0.3461 1.7481
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10425 0.04675 23.62 < 2e-16 ***
## FirstAuthorFemale1 0.01366 0.02354 0.58 0.5620
## LastAuthorFemale1 0.03183 0.02506 1.27 0.2042
## UniqueAuthors2 0.18128 0.02480 7.31 3.3e-13 ***
## UniqueAuthors3 0.19518 0.02600 7.51 7.9e-14 ***
## UniqueAuthors4 0.18821 0.03300 5.70 1.3e-08 ***
## UniqueAuthors5 0.29077 0.03348 8.69 < 2e-16 ***
## Year1997 -0.04219 0.05978 -0.71 0.4804
## Year1998 -0.17541 0.06463 -2.71 0.0067 **
## Year1999 -0.06852 0.05576 -1.23 0.2192
```

```

## Year2000      -0.12849    0.06082    -2.11    0.0347 *
## Year2001      -0.04675    0.06084    -0.77    0.4423
## Year2002      -0.09837    0.06144    -1.60    0.1095
## Year2003      -0.13424    0.05853    -2.29    0.0219 *
## Year2004      -0.17836    0.06010    -2.97    0.0030 **
## Year2005      -0.14998    0.06373    -2.35    0.0187 *
## Year2006      -0.11331    0.06290    -1.80    0.0718 .
## Year2007      -0.12012    0.06251    -1.92    0.0548 .
## Year2008      -0.18236    0.05965    -3.06    0.0023 **
## Year2009      -0.16055    0.05963    -2.69    0.0071 **
## Year2010      -0.05599    0.05826    -0.96    0.3366
## Year2011      -0.14304    0.05812    -2.46    0.0139 *
## Year2012      -0.00573    0.05703    -0.10    0.9200
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0462, Adjusted R-squared:  0.0396
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 279 weights are ~= 1. The remaining 2908 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.203  0.856  0.949   0.896   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.14e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.172 1      1.083
## LastAuthorFemale  1.154 1      1.074
## Year              1.069 16      1.002

```

## Residuals from first and last author



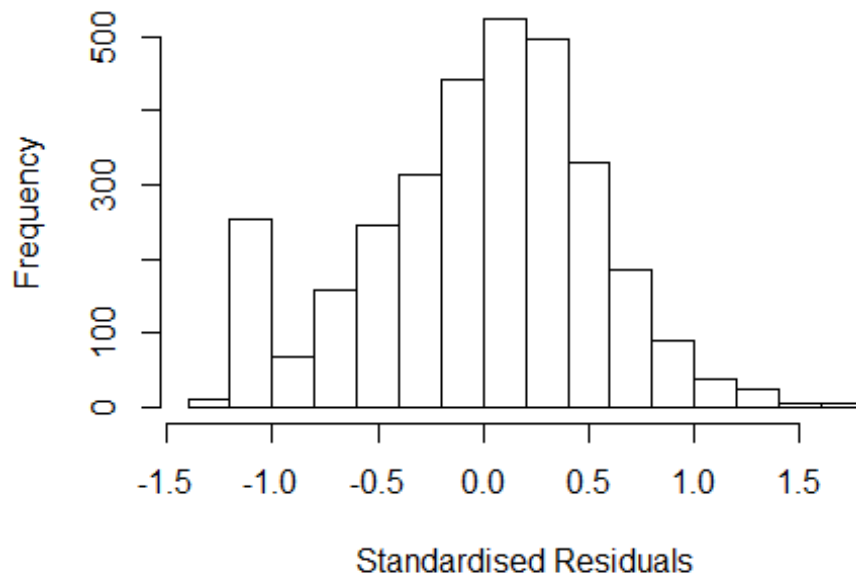
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2329 -0.3415 0.0388 0.3470 1.7088
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20215 0.04529 26.55 <2e-16 ***
## FirstAuthorFemale1 0.02169 0.02418 0.90 0.3697
## LastAuthorFemale1 0.01306 0.02567 0.51 0.6110
## Year1997 -0.04978 0.05929 -0.84 0.4012
## Year1998 -0.16821 0.06465 -2.60 0.0093 **
## Year1999 -0.06452 0.05564 -1.16 0.2463
## Year2000 -0.10787 0.06219 -1.73 0.0829 .
## Year2001 -0.02056 0.06114 -0.34 0.7367
## Year2002 -0.08194 0.06230 -1.32 0.1885
## Year2003 -0.12314 0.05883 -2.09 0.0364 *
## Year2004 -0.14855 0.06072 -2.45 0.0145 *
## Year2005 -0.15440 0.06458 -2.39 0.0169 *
```

```

## Year2006          -0.11160      0.06277    -1.78    0.0755 .
## Year2007          -0.09290      0.06379    -1.46    0.1454
## Year2008          -0.16869      0.06063    -2.78    0.0054 **
## Year2009          -0.15023      0.05995    -2.51    0.0123 *
## Year2010          -0.03116      0.05867    -0.53    0.5954
## Year2011          -0.11633      0.05871    -1.98    0.0476 *
## Year2012          -0.00403      0.05690    -0.07    0.9435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.011, Adjusted R-squared:  0.00541
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 263 weights are ~= 1. The remaining 2924 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.233  0.852  0.948  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.14e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1      1.023
## Year              1.046 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2275 -0.3424 0.0384 0.3497 1.7069
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20414 0.04498 26.77 <2e-16 ***
## FirstAuthorFemale1 0.02786 0.02320 1.20 0.2299
## Year1997 -0.05039 0.05929 -0.85 0.3954
## Year1998 -0.17009 0.06444 -2.64 0.0083 **
## Year1999 -0.06572 0.05559 -1.18 0.2372
## Year2000 -0.10800 0.06219 -1.74 0.0826 .
## Year2001 -0.02100 0.06116 -0.34 0.7314
## Year2002 -0.08247 0.06229 -1.32 0.1856
## Year2003 -0.12379 0.05882 -2.10 0.0354 *
## Year2004 -0.14972 0.06061 -2.47 0.0136 *
## Year2005 -0.15507 0.06453 -2.40 0.0163 *
## Year2006 -0.11271 0.06271 -1.80 0.0724 .
```

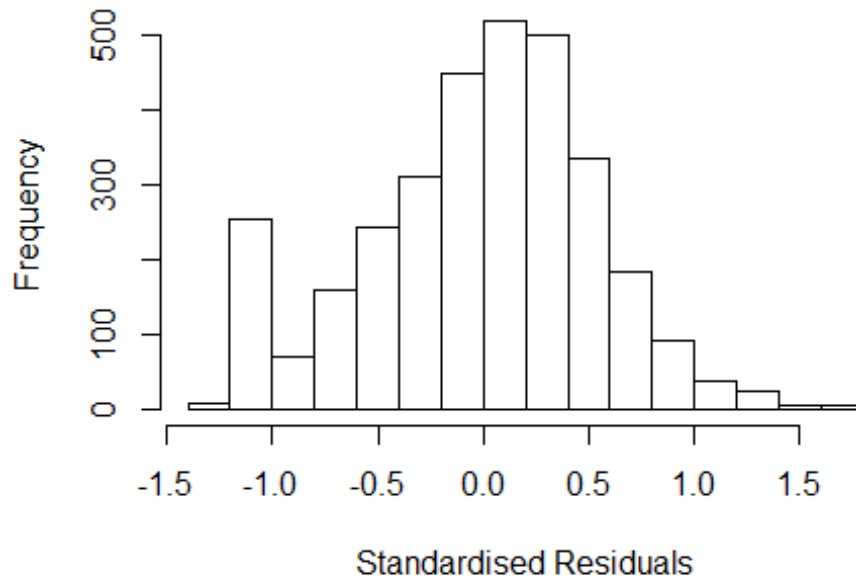


```

## Year2007          -0.09337    0.06374   -1.46    0.1431
## Year2008          -0.16892    0.06069   -2.78    0.0054 **
## Year2009          -0.15020    0.05995   -2.51    0.0123 *
## Year2010          -0.03171    0.05867   -0.54    0.5888
## Year2011          -0.11750    0.05868   -2.00    0.0453 *
## Year2012          -0.00446    0.05692   -0.08    0.9376
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.011, Adjusted R-squared:  0.00565
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 259 weights are ~= 1. The remaining 2928 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.234  0.851  0.948  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.14e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

## Residuals from last author



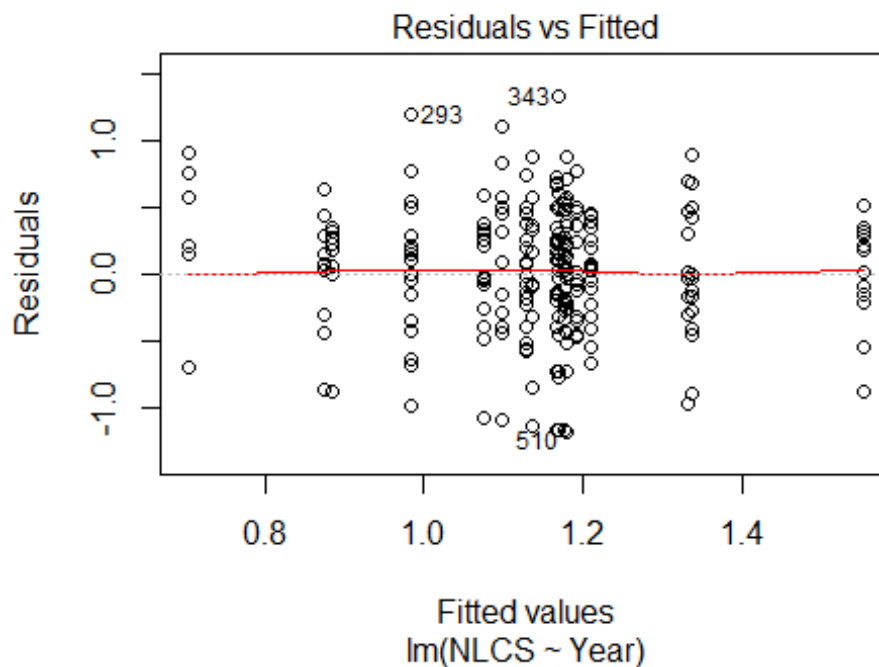
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2279 -0.3410  0.0414  0.3512  1.7072
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20381    0.04518   26.65  <2e-16 ***
## LastAuthorFemale1 0.02413    0.02452    0.98  0.3251
## Year1997       -0.05088    0.05926   -0.86  0.3906
## Year1998       -0.16668    0.06465   -2.58  0.0100 **
## Year1999       -0.06378    0.05563   -1.15  0.2517
## Year2000       -0.10754    0.06222   -1.73  0.0840 .
## Year2001       -0.01992    0.06120   -0.33  0.7449
## Year2002       -0.08103    0.06231   -1.30  0.1935
## Year2003       -0.12270    0.05881   -2.09  0.0370 *
## Year2004       -0.14732    0.06072   -2.43  0.0153 *
## Year2005       -0.15384    0.06462   -2.38  0.0173 *
## Year2006       -0.10922    0.06268   -1.74  0.0815 .
```

```

## Year2007          -0.09219      0.06382    -1.44    0.1487
## Year2008          -0.16715      0.06054    -2.76    0.0058 **
## Year2009          -0.14843      0.05999    -2.47    0.0134 *
## Year2010          -0.02935      0.05858    -0.50    0.6163
## Year2011          -0.11324      0.05850    -1.94    0.0530 .
## Year2012          -0.00133      0.05674    -0.02    0.9813
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00551
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 256 weights are ~= 1. The remaining 2931 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.233  0.852  0.948  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.14e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3187"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1905"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   35   31   36   41   38   50   34   17   34   14   26   16   14   35   25
## 2011 2012
##   19   26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   19   14   17   12   11   21    9   17    8   13   11   10   24   16
## 2011 2012

```

```
## 16 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 18 12 15 11 11 20 9 17 8 13 9 10 21 16
## 2011 2012
## 15 19
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.2
```



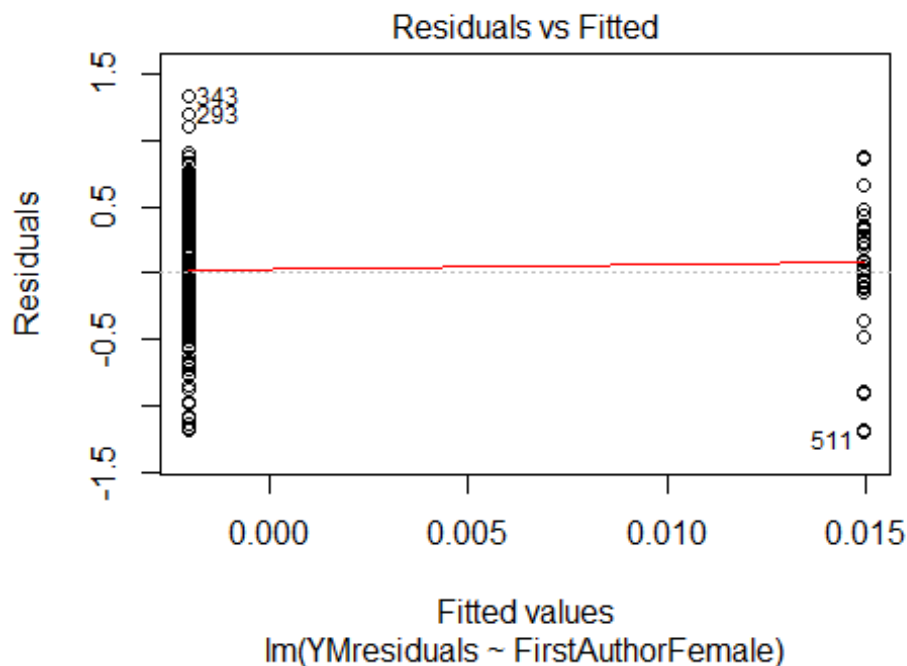
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 1.58489319246111"
## [1] "Male first author team size 2018 geometric mean: 2.19940669555929"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

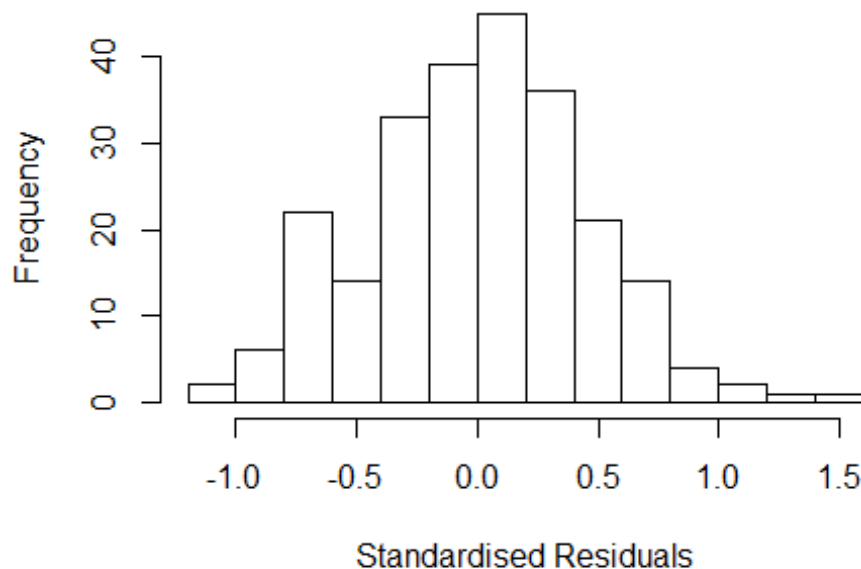
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.43096908110526"
## [1] "Male last author team size 2018 geometric mean: 2.24880159501847"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.052 1      1.432
## LastAuthorFemale  1.376 1      1.173
## UniqueAuthors    5.398 4      1.235
## Year              6.310 16     1.059
```

## Residuals from first and last author and team size



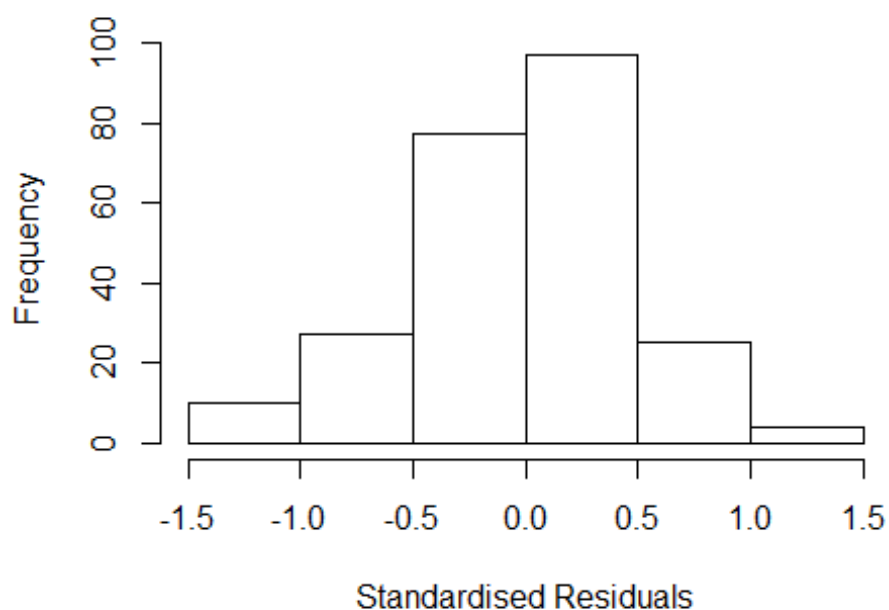
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0064 -0.2550 0.0194 0.2715 1.4772
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0321 0.0946 10.91 < 2e-16 ***
## FirstAuthorFemale1 -0.0919 0.1023 -0.90 0.37019
## LastAuthorFemale1 -0.2713 0.1050 -2.58 0.01046 *
## UniqueAuthors2 0.1707 0.0920 1.86 0.06489 .
## UniqueAuthors3 0.4354 0.0989 4.40 1.7e-05 ***
## UniqueAuthors4 0.3343 0.0931 3.59 0.00041 ***
## UniqueAuthors5 0.4578 0.1249 3.67 0.00031 ***
## Year1997 0.0185 0.1193 0.16 0.87663
## Year1998 -0.1640 0.2113 -0.78 0.43848
## Year1999 0.0845 0.1558 0.54 0.58796
```

```

## Year2000          0.3959      0.1483      2.67  0.00819 **
## Year2001          -0.3974      0.3471     -1.14  0.25349
## Year2002          -0.2402      0.1450     -1.66  0.09923 .
## Year2003          -0.3642      0.1552     -2.35  0.01989 *
## Year2004          -0.0257      0.1548     -0.17  0.86823
## Year2005           0.1005      0.2195      0.46  0.64766
## Year2006           0.0859      0.1787      0.48  0.63143
## Year2007          -0.0969      0.1791     -0.54  0.58923
## Year2008          -0.1732      0.1188     -1.46  0.14610
## Year2009          -0.0600      0.1489     -0.40  0.68712
## Year2010          -0.1368      0.1283     -1.07  0.28762
## Year2011          -0.0316      0.1462     -0.22  0.82887
## Year2012          -0.0299      0.1317     -0.23  0.82087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.244, Adjusted R-squared:  0.167
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 221 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.202  0.849  0.959  0.899  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.17e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.848 1          1.359
## LastAuthorFemale  1.360 1          1.166
## Year              2.273 16          1.026

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2281 -0.3075 0.0339 0.2816 1.3383
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15854 0.09399 12.33 <2e-16 ***
## FirstAuthorFemale1 0.05635 0.10971 0.51 0.6081
## LastAuthorFemale1 -0.24404 0.12471 -1.96 0.0516 .
## Year1997 0.04520 0.12069 0.37 0.7084
## Year1998 -0.13958 0.25930 -0.54 0.5909
## Year1999 0.08839 0.15413 0.57 0.5669
## Year2000 0.42723 0.15510 2.75 0.0064 **
## Year2001 -0.48622 0.36369 -1.34 0.1826
## Year2002 -0.18873 0.16193 -1.17 0.2451
## Year2003 -0.27570 0.18530 -1.49 0.1382
## Year2004 0.00713 0.17009 0.04 0.9666
## Year2005 0.20737 0.19882 1.04 0.2981
```

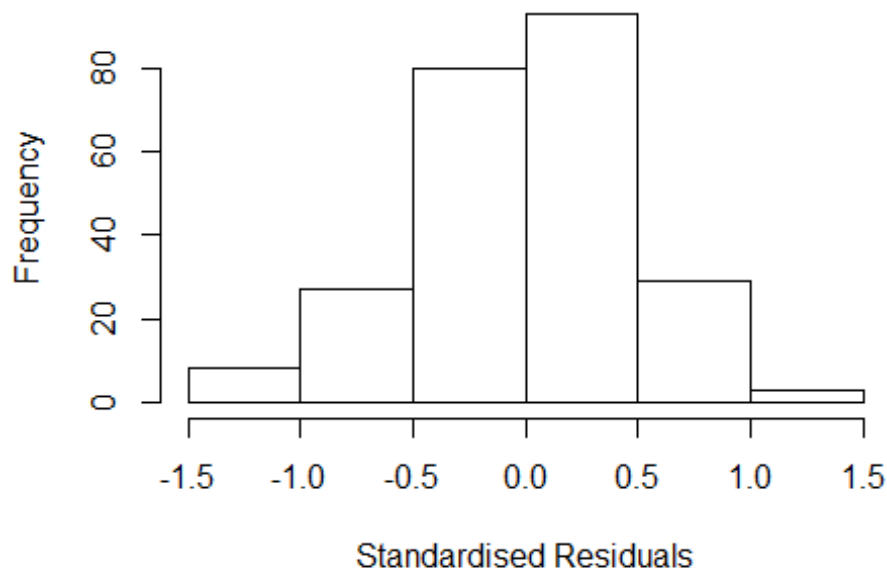


```

## Year2006          0.23863      0.18052      1.32      0.1876
## Year2007          -0.02980      0.25959     -0.11      0.9087
## Year2008          -0.14455      0.14933     -0.97      0.3341
## Year2009           0.01226      0.14285      0.09      0.9317
## Year2010          -0.04564      0.13600     -0.34      0.7375
## Year2011           0.04883      0.13427      0.36      0.7165
## Year2012           0.06956      0.12544      0.55      0.5797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.138, Adjusted R-squared:  0.0676
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 219 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.353  0.862  0.955  0.896  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.823 1      1.350
## Year              1.823 16      1.019

```

## Residuals from first author

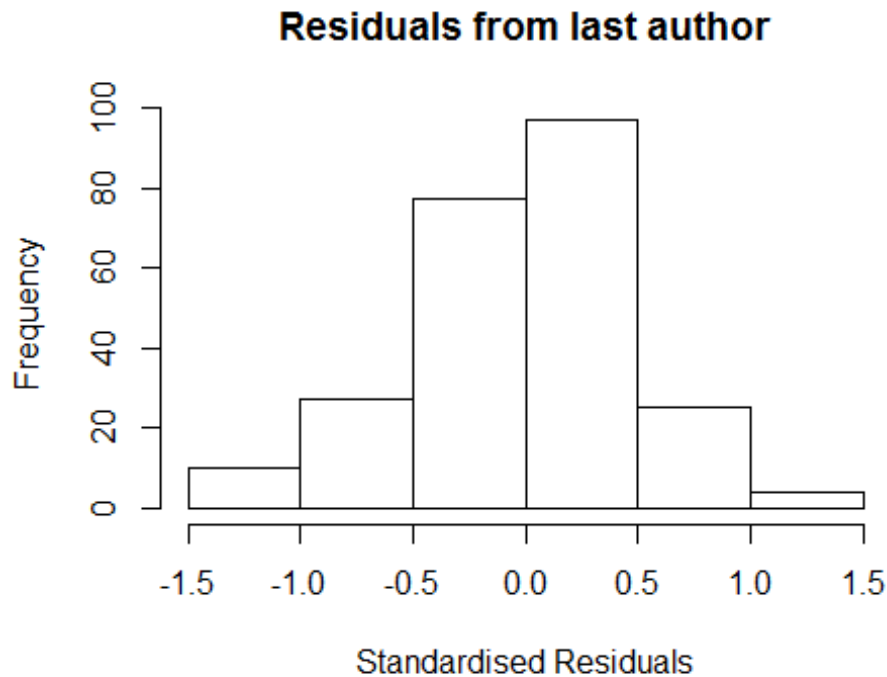


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2621 -0.2908 0.0147 0.2776 1.3384
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16074 0.09469 12.26 <2e-16 ***
## FirstAuthorFemale1 0.02100 0.12434 0.17 0.8660
## Year1997 0.03180 0.12053 0.26 0.7922
## Year1998 -0.17459 0.24335 -0.72 0.4739
## Year1999 0.08040 0.15524 0.52 0.6051
## Year2000 0.42803 0.15587 2.75 0.0065 **
## Year2001 -0.48690 0.35624 -1.37 0.1731
## Year2002 -0.20100 0.15861 -1.27 0.2064
## Year2003 -0.26872 0.18685 -1.44 0.1518
## Year2004 0.00491 0.17035 0.03 0.9770
## Year2005 0.20413 0.19909 1.03 0.3063
## Year2006 0.16232 0.17332 0.94 0.3500
```

```

## Year2007          -0.03380    0.25869   -0.13    0.8962
## Year2008          -0.21341    0.17037   -1.25    0.2117
## Year2009          -0.00218    0.14483   -0.02    0.9880
## Year2010          -0.05966    0.13678   -0.44    0.6631
## Year2011           0.03895    0.13768    0.28    0.7775
## Year2012           0.04208    0.12411    0.34    0.7349
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.124, Adjusted R-squared:  0.057
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.366  0.854  0.953  0.895  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.224 1          1.106
## Year              1.224 16          1.006

```



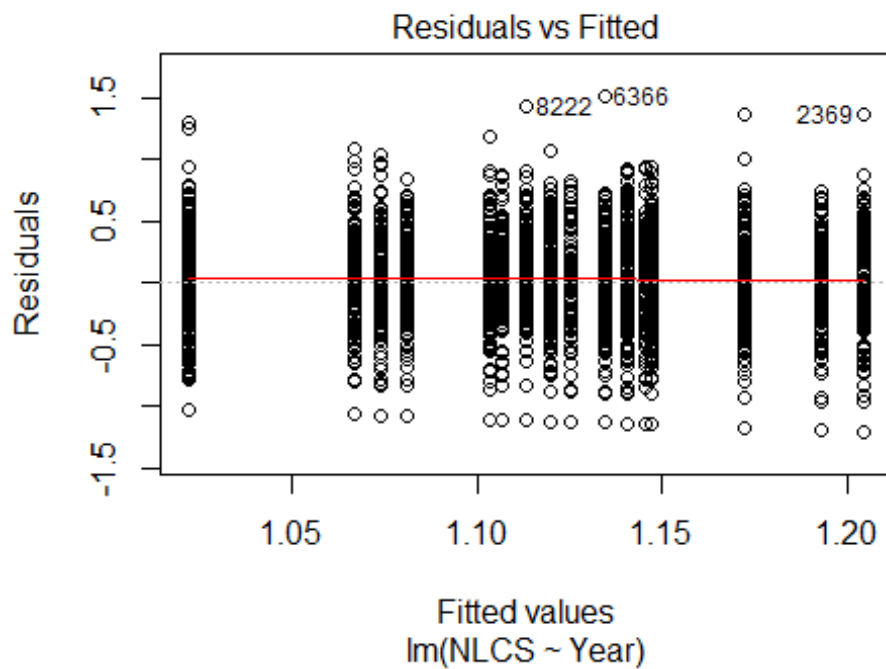
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2370 -0.3123 0.0332 0.2832 1.3383
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16181 0.09539 12.18 <2e-16 ***
## LastAuthorFemale1 -0.22960 0.12728 -1.80 0.0726 .
## Year1997 0.04797 0.12216 0.39 0.6949
## Year1998 -0.14477 0.26119 -0.55 0.5800
## Year1999 0.10072 0.15648 0.64 0.5205
## Year2000 0.43113 0.15656 2.75 0.0064 **
## Year2001 -0.49105 0.37180 -1.32 0.1879
## Year2002 -0.18598 0.16383 -1.14 0.2575
## Year2003 -0.26240 0.18398 -1.43 0.1552
## Year2004 0.00391 0.17101 0.02 0.9818
## Year2005 0.20513 0.19963 1.03 0.3053
## Year2006 0.22897 0.17624 1.30 0.1952
```

```

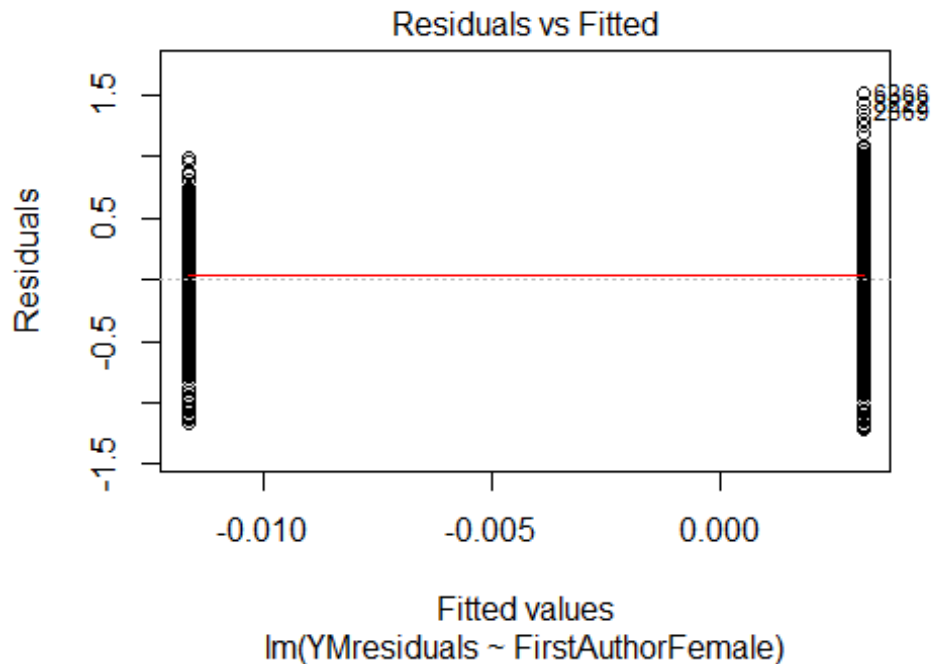
## Year2007          -0.03127      0.26135    -0.12    0.9049
## Year2008          -0.14763      0.14550    -1.01    0.3114
## Year2009           0.01235      0.14406     0.09    0.9318
## Year2010          -0.03666      0.13531    -0.27    0.7867
## Year2011           0.05637      0.13340     0.42    0.6730
## Year2012           0.07522      0.12504     0.60    0.5481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.138, Adjusted R-squared:  0.0719
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 220 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.340  0.857  0.952  0.894  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.17e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 240"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1906"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 335 363 351 321 347 328 341 329 244 269 313 270 308 273 264
## 2011 2012
## 286 295
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 136 154 154 140 123 104 170 134 106 161 185 155 195 156 149
## 2011 2012

```

```
## 158 162
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 122 141 139 124 115 86 144 122 99 133 164 136 172 135 138
## 2011 2012
## 142 150
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```

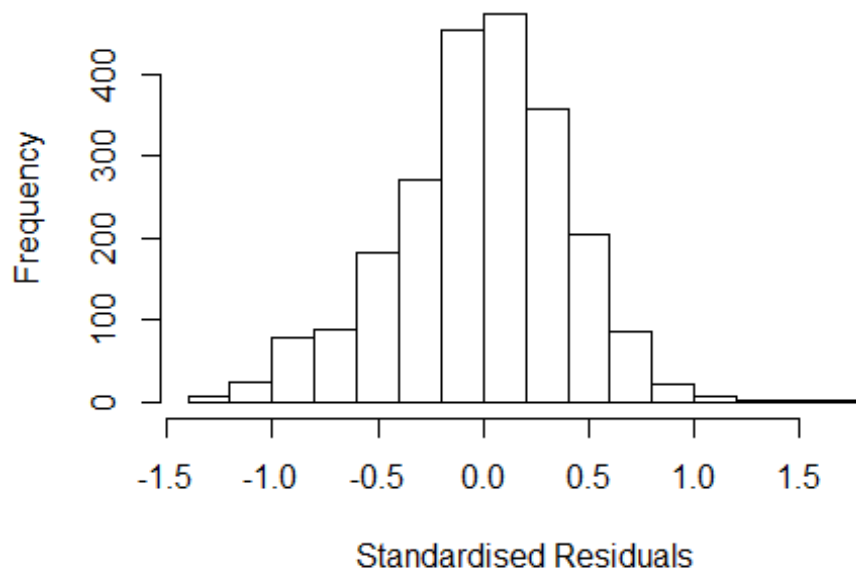


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.4, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 3.21346148678587"
## [1] "Male first author team size 2018 geometric mean: 2.81210421144346"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.61521523145389"
## [1] "Male last author team size 2018 geometric mean: 2.96992571813118"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 850, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1      1.028
## LastAuthorFemale  1.046 1      1.023
## UniqueAuthors    1.176 4      1.021
## Year             1.224 16      1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2888 -0.2490  0.0107  0.2632  1.7092
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.97079    0.04537   21.40 < 2e-16 ***
## FirstAuthorFemale1 -0.03523    0.02095   -1.68  0.093 .
## LastAuthorFemale1  0.00790    0.02810    0.28  0.779
## UniqueAuthors2     0.20443    0.02933    6.97 4.1e-12 ***
## UniqueAuthors3     0.24170    0.02945    8.21 3.8e-16 ***
## UniqueAuthors4     0.27375    0.03136    8.73 < 2e-16 ***
## UniqueAuthors5     0.34956    0.03289   10.63 < 2e-16 ***
## Year1997          0.04667    0.05391    0.87  0.387
## Year1998         -0.00539    0.05529   -0.10  0.922
## Year1999          0.07044    0.05216    1.35  0.177
```

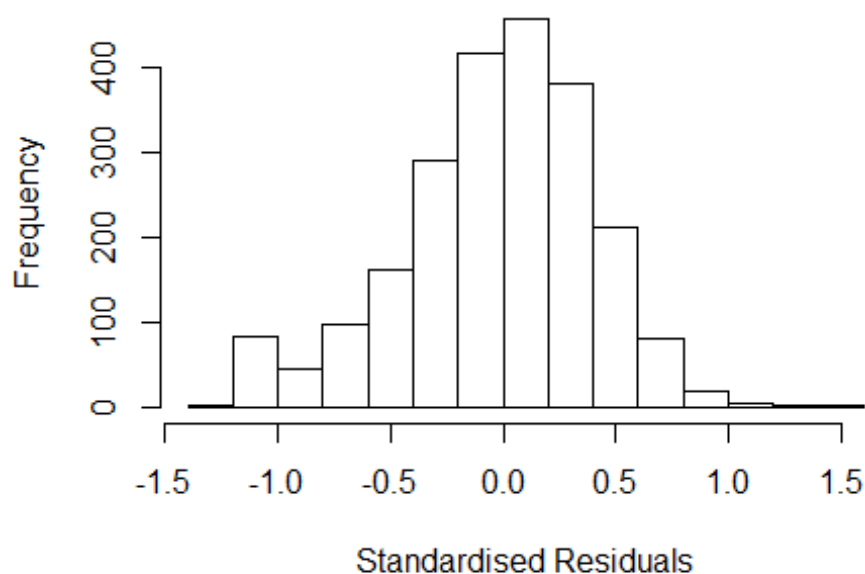


```

## Year2000      0.06842    0.05736    1.19    0.233
## Year2001     -0.01936    0.06146   -0.32    0.753
## Year2002      0.01087    0.05615    0.19    0.847
## Year2003     -0.02831    0.05495   -0.52    0.607
## Year2004     -0.02051    0.05620   -0.36    0.715
## Year2005     -0.04597    0.05740   -0.80    0.423
## Year2006     -0.01093    0.05010   -0.22    0.827
## Year2007     -0.12365    0.05682   -2.18    0.030 *
## Year2008     -0.03295    0.05082   -0.65    0.517
## Year2009     -0.04602    0.05608   -0.82    0.412
## Year2010     -0.11971    0.05365   -2.23    0.026 *
## Year2011     -0.09241    0.05309   -1.74    0.082 .
## Year2012     -0.03934    0.05170   -0.76    0.447
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0843, Adjusted R-squared:  0.0753
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 192 weights are ~= 1. The remaining 2070 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0107 0.8590 0.9500 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## LastAuthorFemale  1.052 1      1.025
## Year              1.077 16      1.002

```

## Residuals from first and last author



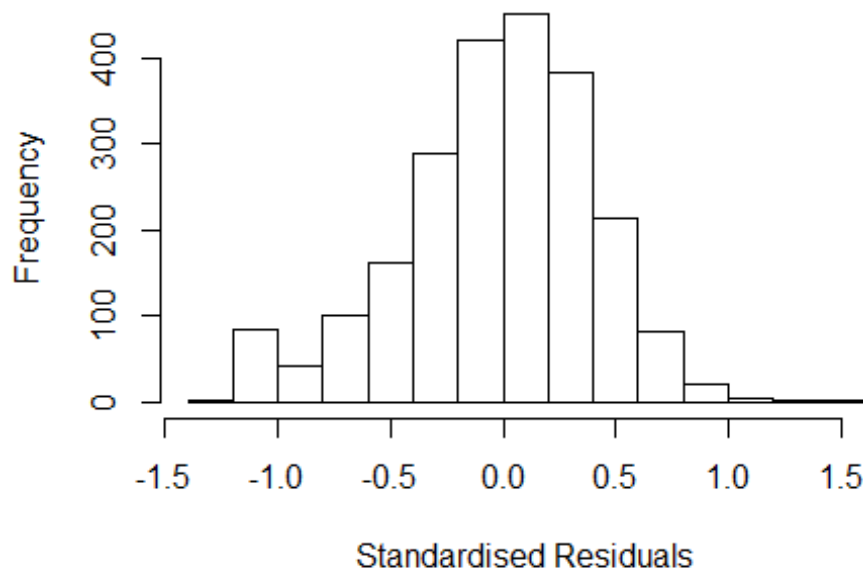
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2415 -0.2612 0.0115 0.2687 1.5024
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.134464 0.044228 25.65 <2e-16 ***
## FirstAuthorFemale1 -0.007168 0.020898 -0.34 0.732
## LastAuthorFemale1 0.013077 0.028296 0.46 0.644
## Year1997 0.038557 0.057316 0.67 0.501
## Year1998 0.009830 0.057454 0.17 0.864
## Year1999 0.070895 0.054454 1.30 0.193
## Year2000 0.094004 0.060793 1.55 0.122
## Year2001 0.011131 0.065160 0.17 0.864
## Year2002 0.037438 0.058158 0.64 0.520
## Year2003 -0.000163 0.056326 0.00 0.998
## Year2004 0.011914 0.059026 0.20 0.840
## Year2005 -0.027825 0.060334 -0.46 0.645
```

```

## Year2006          0.021053    0.052129    0.40    0.686
## Year2007          -0.088035    0.060554   -1.45    0.146
## Year2008          0.010179    0.053946    0.19    0.850
## Year2009          -0.019024    0.059388   -0.32    0.749
## Year2010          -0.096021    0.057265   -1.68    0.094 .
## Year2011          -0.041346    0.055732   -0.74    0.458
## Year2012          0.013097    0.053779    0.24    0.808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0128, Adjusted R-squared:  0.00488
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 217 weights are ~= 1. The remaining 2045 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.861  0.949  0.890  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.034 1      1.017
## Year              1.034 16      1.001

```

## Residuals from first author



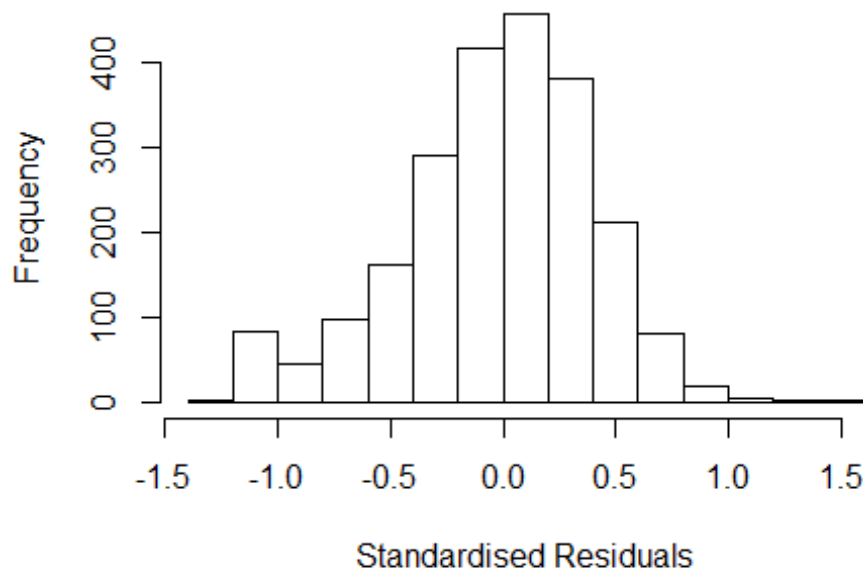
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2294 -0.2595 0.0112 0.2683 1.5016
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.135321 0.044274 25.64 <2e-16 ***
## FirstAuthorFemale1 -0.005568 0.020838 -0.27 0.789
## Year1997 0.039363 0.057382 0.69 0.493
## Year1998 0.010163 0.057584 0.18 0.860
## Year1999 0.070450 0.054502 1.29 0.196
## Year2000 0.094097 0.060824 1.55 0.122
## Year2001 0.010833 0.065114 0.17 0.868
## Year2002 0.037699 0.058176 0.65 0.517
## Year2003 -0.000316 0.056396 -0.01 0.996
## Year2004 0.011216 0.059101 0.19 0.850
## Year2005 -0.027763 0.060371 -0.46 0.646
## Year2006 0.022223 0.052072 0.43 0.670
```

```

## Year2007          -0.087421    0.060579   -1.44    0.149
## Year2008          0.010036    0.053981    0.19    0.853
## Year2009         -0.018166    0.059425   -0.31    0.760
## Year2010         -0.094844    0.057186   -1.66    0.097 .
## Year2011         -0.041157    0.055832   -0.74    0.461
## Year2012          0.013287    0.053866    0.25    0.805
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0127, Adjusted R-squared:  0.00519
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 213 weights are ~= 1. The remaining 2049 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.112  0.861  0.950  0.890  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.042 1          1.021
## Year              1.042 16          1.001

```

## Residuals from last author



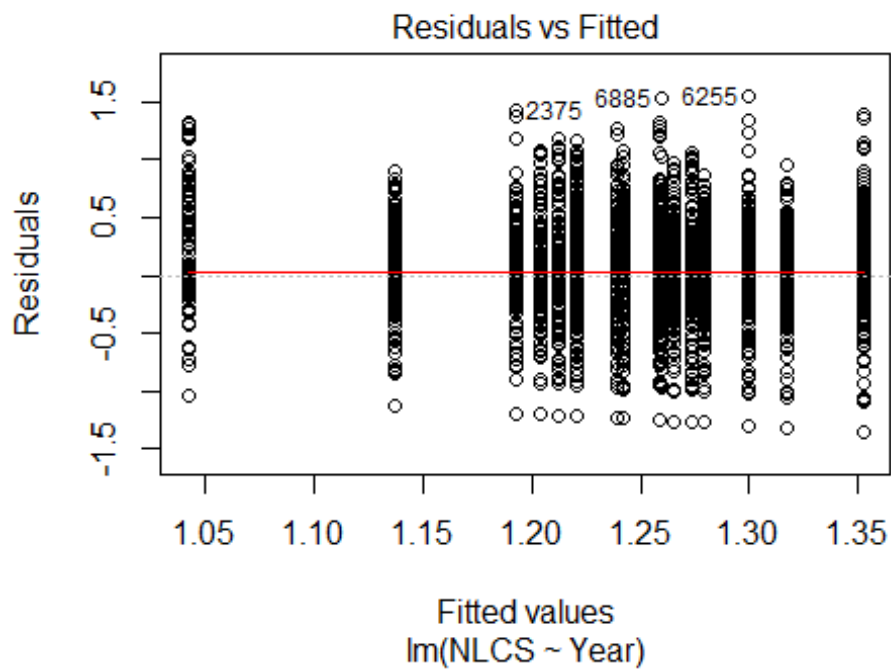
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2387 -0.2609  0.0128  0.2690  1.5043
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13370    0.04406   25.73  <2e-16 ***
## LastAuthorFemale1 0.01168    0.02821    0.41   0.679
## Year1997        0.03878    0.05728    0.68   0.498
## Year1998        0.00917    0.05744    0.16   0.873
## Year1999        0.07061    0.05445    1.30   0.195
## Year2000        0.09334    0.06088    1.53   0.125
## Year2001        0.01048    0.06509    0.16   0.872
## Year2002        0.03732    0.05815    0.64   0.521
## Year2003       -0.00118    0.05628   -0.02   0.983
## Year2004        0.01099    0.05900    0.19   0.852
## Year2005       -0.02856    0.06034   -0.47   0.636
## Year2006        0.02062    0.05215    0.40   0.693
```

```

## Year2007          -0.08858      0.06051    -1.46      0.143
## Year2008           0.00905      0.05381      0.17      0.867
## Year2009          -0.01991      0.05939     -0.34      0.737
## Year2010          -0.09654      0.05725     -1.69      0.092 .
## Year2011          -0.04241      0.05569     -0.76      0.446
## Year2012           0.01180      0.05367      0.22      0.826
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0128, Adjusted R-squared:  0.00529
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2053 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.862  0.949  0.890  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2262"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1907"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  379  378  379  416  435  377  397  272  273  268  306  254  256  302  291
## 2011 2012
##  289  287
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  153  138  157  179  122  100  181  129  130  146  176  138  168  185  162
## 2011 2012

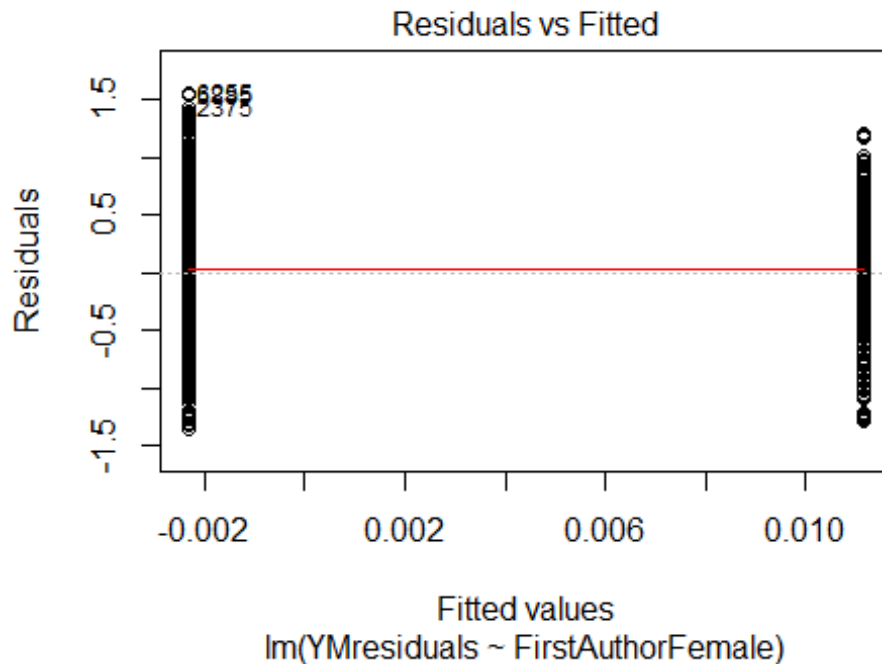
```

```
## 187 176
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 140 126 145 171 112 85 152 119 125 124 151 121 147 166 152
## 2011 2012
## 172 157
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 93, df = 16, p-value = 9e-13
```



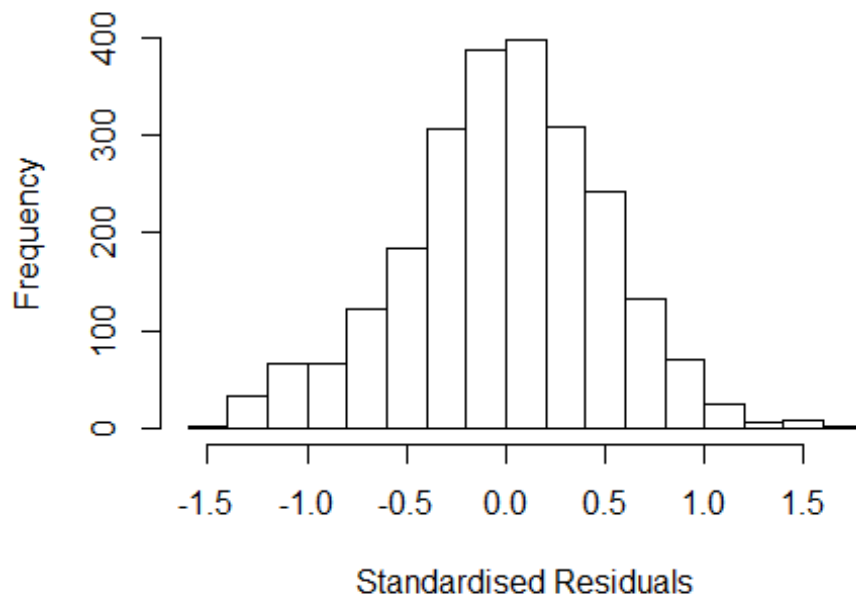
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.2, df = 1, p-value = 0.04
```





```
## [1] "Female first author team size 2018 geometric mean: 3.38569068180038"
## [1] "Male first author team size 2018 geometric mean: 2.70260628928169"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4700, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.29722947114067"
## [1] "Male last author team size 2018 geometric mean: 2.75829843424107"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3700, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1          1.021
## LastAuthorFemale  1.079 1          1.039
## UniqueAuthors     1.203 4          1.023
## Year              1.289 16         1.008
```

## Residuals from first and last author and team size



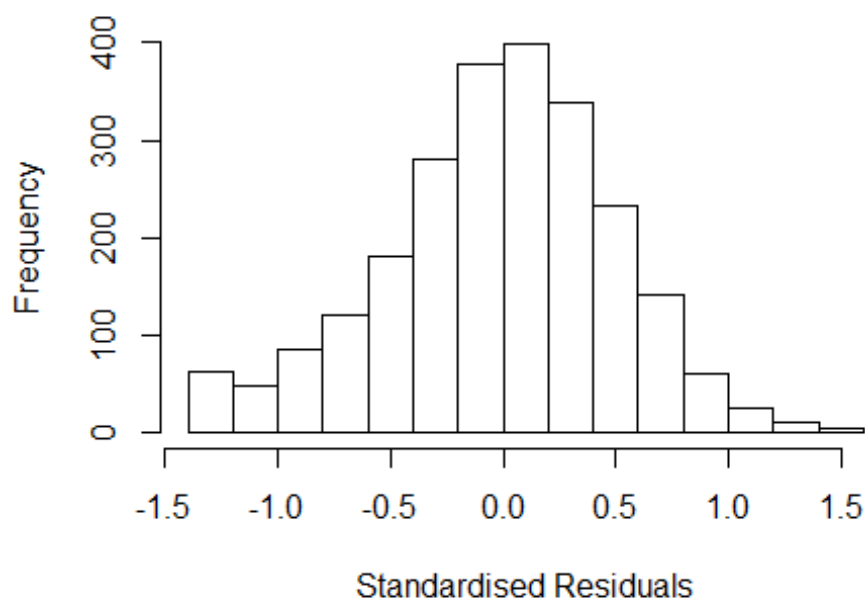
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42558 -0.31310  0.00456  0.32196  1.71628
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17294    0.05059   23.19  < 2e-16 ***
## FirstAuthorFemale1  0.00417    0.02691    0.15  0.8770
## LastAuthorFemale1 -0.04243    0.03448   -1.23  0.2187
## UniqueAuthors2     0.17397    0.03007    5.78 8.2e-09 ***
## UniqueAuthors3     0.22130    0.03126    7.08 1.9e-12 ***
## UniqueAuthors4     0.27236    0.03754    7.25 5.5e-13 ***
## UniqueAuthors5     0.28504    0.03734    7.63 3.3e-14 ***
## Year1997          -0.02632    0.07559   -0.35  0.7277
## Year1998          -0.06369    0.07122   -0.89  0.3713
## Year1999          -0.01972    0.06378   -0.31  0.7572
```

```

## Year2000      -0.12956    0.06564   -1.97    0.0485 *
## Year2001      -0.24022    0.11432   -2.10    0.0357 *
## Year2002       0.04184    0.06453    0.65    0.5168
## Year2003      -0.09328    0.06776   -1.38    0.1687
## Year2004      -0.09506    0.06487   -1.47    0.1429
## Year2005      -0.11487    0.06725   -1.71    0.0877 .
## Year2006      -0.03853    0.06298   -0.61    0.5407
## Year2007       0.01002    0.05969    0.17    0.8667
## Year2008      -0.02927    0.06151   -0.48    0.6342
## Year2009      -0.04093    0.05843   -0.70    0.4837
## Year2010      -0.09622    0.05914   -1.63    0.1038
## Year2011      -0.18379    0.05853   -3.14    0.0017 **
## Year2012      -0.09280    0.06112   -1.52    0.1291
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0566, Adjusted R-squared:  0.0478
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 226 weights are ~= 1. The remaining 2139 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.171  0.861  0.951  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1      1.016
## LastAuthorFemale  1.083 1      1.041
## Year              1.098 16      1.003

```

## Residuals from first and last author



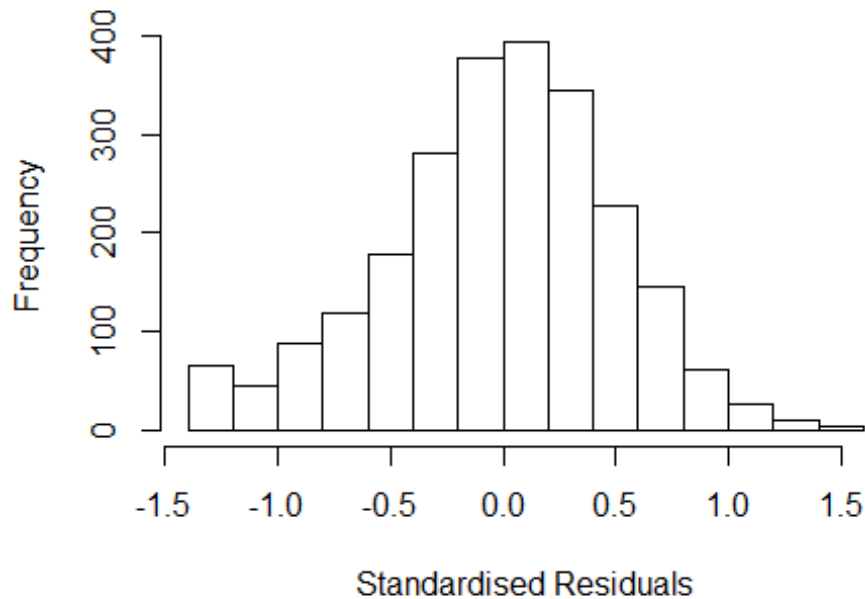
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3731 -0.3250  0.0147  0.3347  1.5398
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28066    0.04999   25.62  <2e-16 ***
## FirstAuthorFemale1  0.03333    0.02685    1.24   0.215
## LastAuthorFemale1 -0.03628    0.03492   -1.04   0.299
## Year1997         -0.00417    0.07784   -0.05   0.957
## Year1998         -0.05343    0.07328   -0.73   0.466
## Year1999         -0.00132    0.06628   -0.02   0.984
## Year2000         -0.09970    0.06744   -1.48   0.139
## Year2001         -0.20063    0.11455   -1.75   0.080 .
## Year2002          0.09239    0.06620    1.40   0.163
## Year2003         -0.06003    0.06995   -0.86   0.391
## Year2004         -0.05299    0.06729   -0.79   0.431
## Year2005         -0.08559    0.07029   -1.22   0.223
```

```

## Year2006          0.00274    0.06450    0.04    0.966
## Year2007          0.04840    0.06191    0.78    0.434
## Year2008          0.02864    0.06272    0.46    0.648
## Year2009          0.00976    0.05960    0.16    0.870
## Year2010         -0.02742    0.05977   -0.46    0.647
## Year2011         -0.11694    0.05994   -1.95    0.051 .
## Year2012         -0.02192    0.06235   -0.35    0.725
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.0158, Adjusted R-squared:  0.00821
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 199 weights are ~= 1. The remaining 2166 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.292  0.862  0.950  0.897  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year              1.025 16      1.001

```

## Residuals from first author



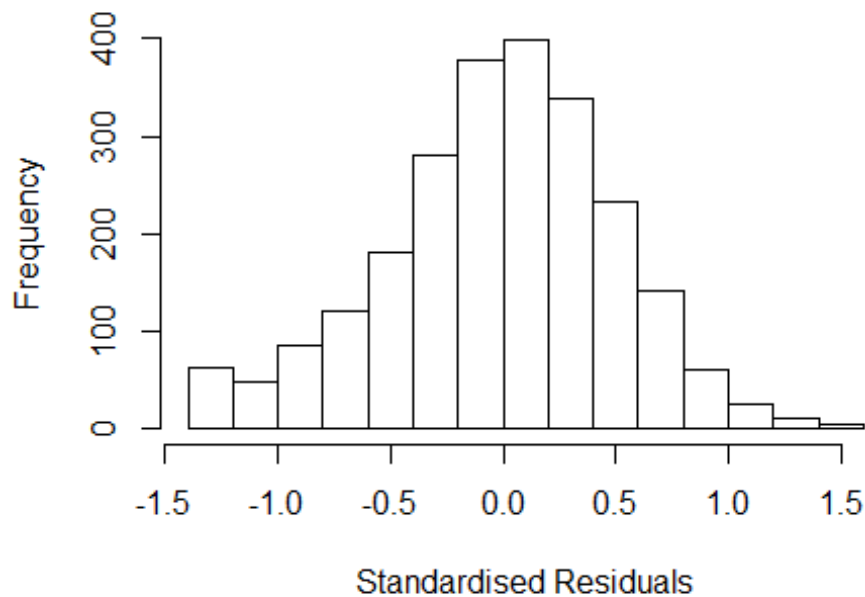
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3692 -0.3228 0.0159 0.3319 1.5442
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.276377 0.049549 25.76 <2e-16 ***
## FirstAuthorFemale1 0.027247 0.027033 1.01 0.314
## Year1997 -0.005072 0.077674 -0.07 0.948
## Year1998 -0.052407 0.073095 -0.72 0.473
## Year1999 0.000601 0.066162 0.01 0.993
## Year2000 -0.100209 0.067442 -1.49 0.137
## Year2001 -0.201567 0.114959 -1.75 0.080 .
## Year2002 0.092841 0.066137 1.40 0.161
## Year2003 -0.057269 0.069579 -0.82 0.411
## Year2004 -0.050956 0.067082 -0.76 0.448
## Year2005 -0.082061 0.069836 -1.18 0.240
## Year2006 0.002835 0.064415 0.04 0.965
```

```

## Year2007          0.046749    0.061823    0.76    0.450
## Year2008          0.030577    0.062622    0.49    0.625
## Year2009          0.011613    0.059405    0.20    0.845
## Year2010         -0.027622    0.059672   -0.46    0.643
## Year2011         -0.114320    0.059735   -1.91    0.056 .
## Year2012         -0.020952    0.062210   -0.34    0.736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.00803
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2156 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.861  0.949  0.896  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.074 1          1.037
## Year            1.074 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3791 -0.3264 0.0146 0.3327 1.5336
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28483 0.04989 25.75 <2e-16 ***
## LastAuthorFemale1 -0.02861 0.03479 -0.82 0.411
## Year1997 -0.00386 0.07798 -0.05 0.961
## Year1998 -0.05400 0.07334 -0.74 0.462
## Year1999 -0.00289 0.06635 -0.04 0.965
## Year2000 -0.10169 0.06745 -1.51 0.132
## Year2001 -0.19829 0.11521 -1.72 0.085 .
## Year2002 0.09423 0.06628 1.42 0.155
## Year2003 -0.05942 0.06995 -0.85 0.396
## Year2004 -0.05488 0.06728 -0.82 0.415
## Year2005 -0.08487 0.07043 -1.21 0.228
## Year2006 0.00352 0.06454 0.05 0.957
```

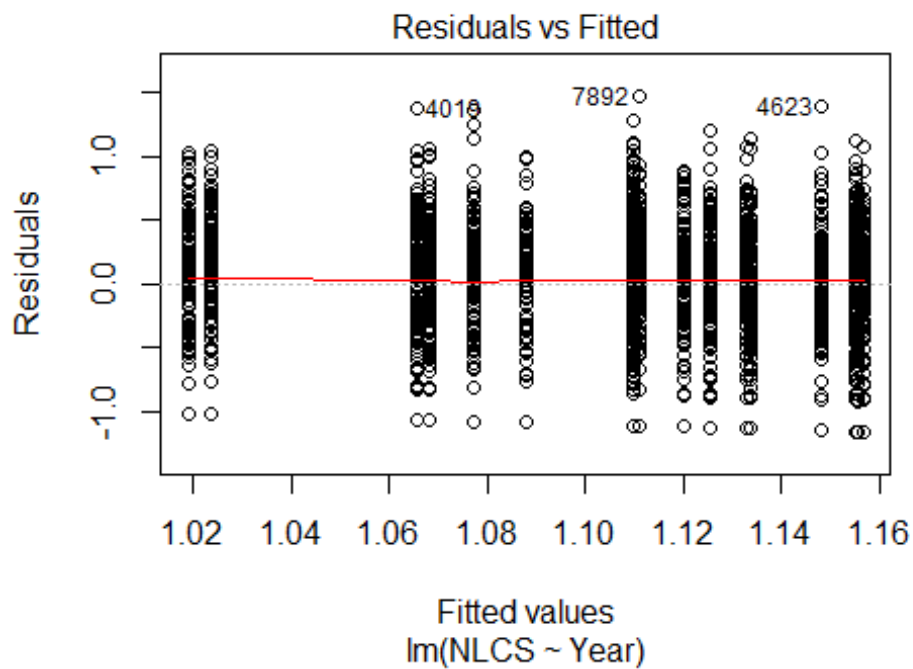


```

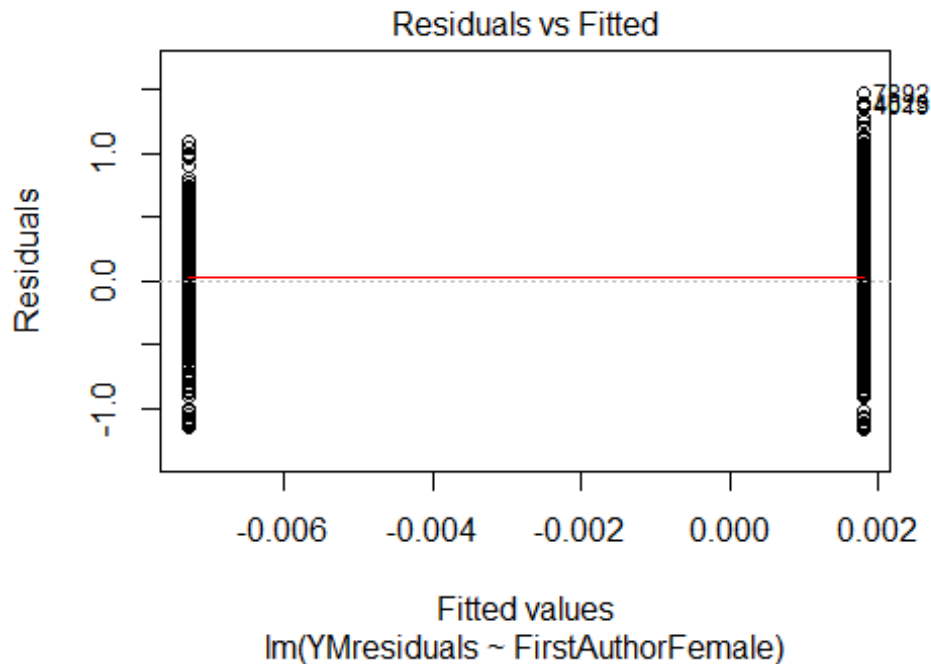
## Year2007      0.04869      0.06190      0.79      0.432
## Year2008      0.03059      0.06292      0.49      0.627
## Year2009      0.01288      0.05972      0.22      0.829
## Year2010     -0.02543      0.05978     -0.43      0.671
## Year2011     -0.11568      0.06001     -1.93      0.054 .
## Year2012     -0.01959      0.06231     -0.31      0.753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.0081
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2165 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.296  0.862  0.949  0.897  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2365"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1908"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 301 314 293 285 312 280 303 290 264 239 288 262 271 272 266
## 2011 2012
## 288 275
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 121 144 137 134 111 91 167 144 124 149 168 152 166 145 133
## 2011 2012

```

```
## 157 153
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 134 123 124 100 80 140 126 116 115 145 128 149 130 121
## 2011 2012
## 140 136
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 16, df = 16, p-value = 0.5
```

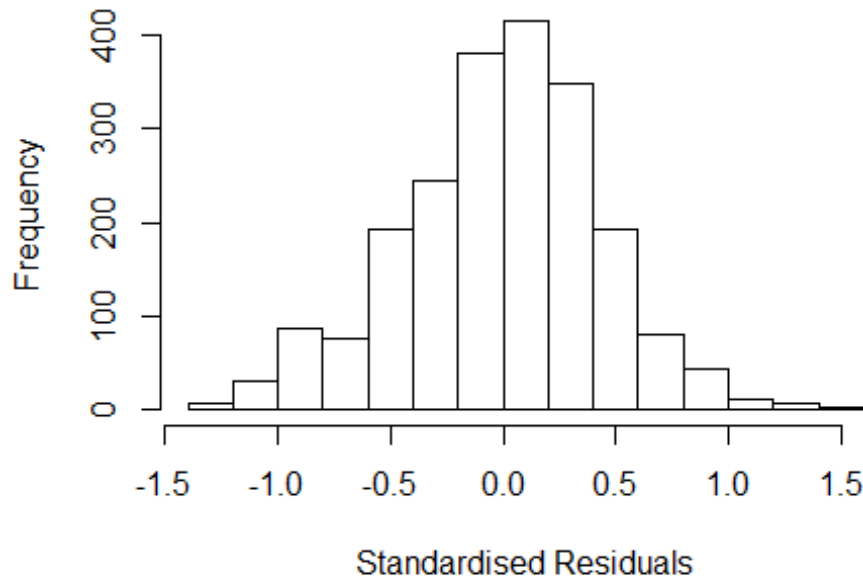


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.4, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 3.23151632748721"
## [1] "Male first author team size 2018 geometric mean: 2.78997739294562"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.94843711059142"
## [1] "Male last author team size 2018 geometric mean: 2.87538625814359"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 850, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1      1.032
## LastAuthorFemale  1.041 1      1.020
## UniqueAuthors    1.240 4      1.027
## Year             1.283 16      1.008
```

## Residuals from first and last author and team size



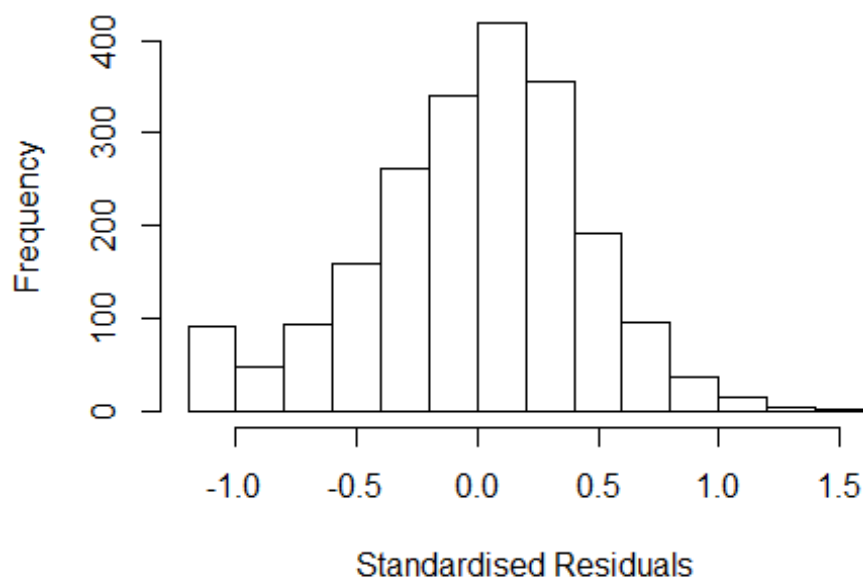
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2605 -0.2754 0.0216 0.2727 1.4600
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0197 0.0515 19.81 < 2e-16 ***
## FirstAuthorFemale1 -0.0311 0.0236 -1.32 0.18827
## LastAuthorFemale1 -0.0183 0.0308 -0.59 0.55331
## UniqueAuthors2 0.2248 0.0283 7.93 3.5e-15 ***
## UniqueAuthors3 0.2713 0.0308 8.81 < 2e-16 ***
## UniqueAuthors4 0.2869 0.0340 8.44 < 2e-16 ***
## UniqueAuthors5 0.4392 0.0357 12.30 < 2e-16 ***
## Year1997 -0.0368 0.0631 -0.58 0.55965
## Year1998 -0.0488 0.0649 -0.75 0.45148
## Year1999 -0.0229 0.0592 -0.39 0.69814
```

```

## Year2000          -0.0122      0.0648   -0.19  0.85081
## Year2001          -0.1130      0.0705   -1.60  0.10918
## Year2002          -0.1476      0.0634   -2.33  0.02008 *
## Year2003          -0.0331      0.0602   -0.55  0.58200
## Year2004          -0.0968      0.0619   -1.56  0.11815
## Year2005          -0.1438      0.0639   -2.25  0.02467 *
## Year2006          -0.1101      0.0568   -1.94  0.05269 .
## Year2007          -0.2184      0.0623   -3.51  0.00046 ***
## Year2008          -0.1030      0.0577   -1.78  0.07452 .
## Year2009          -0.1761      0.0643   -2.74  0.00620 **
## Year2010          -0.1567      0.0602   -2.60  0.00930 **
## Year2011          -0.1351      0.0608   -2.22  0.02628 *
## Year2012          -0.1470      0.0606   -2.43  0.01537 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.0977, Adjusted R-squared:  0.0883
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 181 weights are ~= 1. The remaining 1934 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.173  0.862  0.949  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.73e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.040 1          1.020
## LastAuthorFemale  1.044 1          1.022
## Year              1.067 16          1.002

```

## Residuals from first and last author



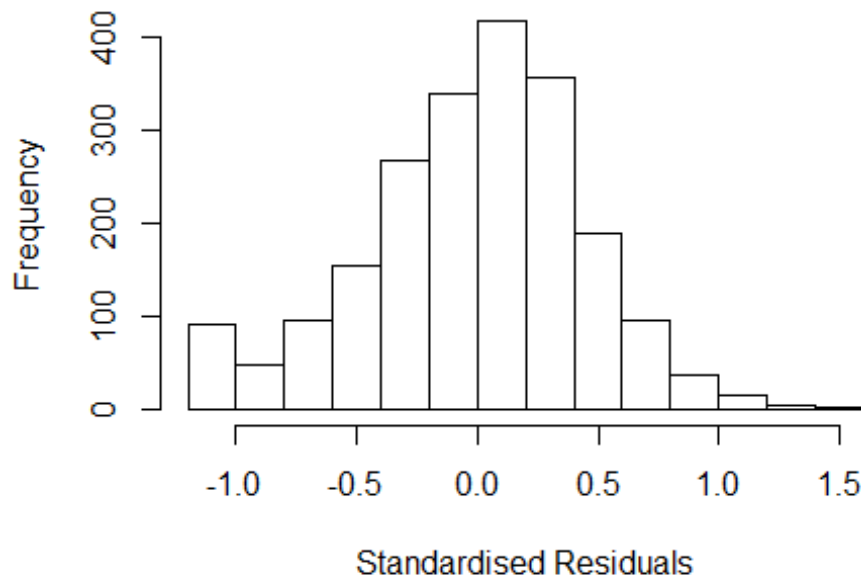
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1959 -0.2844  0.0252  0.2787  1.4622
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18479    0.04947   23.95  <2e-16 ***
## FirstAuthorFemale1 -0.00272    0.02355   -0.12    0.908
## LastAuthorFemale1 -0.01928    0.03087   -0.62    0.532
## Year1997        -0.05777    0.06520   -0.89    0.376
## Year1998        -0.03384    0.06447   -0.52    0.600
## Year1999        -0.02101    0.06137   -0.34    0.732
## Year2000         0.01114    0.06788    0.16    0.870
## Year2001        -0.08099    0.07256   -1.12    0.264
## Year2002        -0.11164    0.06479   -1.72    0.085 .
## Year2003         0.00237    0.06099    0.04    0.969
## Year2004        -0.05733    0.06355   -0.90    0.367
## Year2005        -0.11205    0.06666   -1.68    0.093 .
```

```

## Year2006      -0.06405    0.05780   -1.11    0.268
## Year2007      -0.16456    0.06618   -2.49    0.013 *
## Year2008      -0.04745    0.06090   -0.78    0.436
## Year2009      -0.13729    0.06701   -2.05    0.041 *
## Year2010      -0.11249    0.06248   -1.80    0.072 .
## Year2011      -0.07794    0.06357   -1.23    0.220
## Year2012      -0.06703    0.06205   -1.08    0.280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.00326
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1941 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.203  0.859  0.950  0.893  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1      1.015
## Year      1.031 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1945 -0.2881 0.0253 0.2781 1.4637
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18284 0.04916 24.06 <2e-16 ***
## FirstAuthorFemale1 -0.00557 0.02358 -0.24 0.813
## Year1997 -0.05834 0.06513 -0.90 0.371
## Year1998 -0.03358 0.06430 -0.52 0.602
## Year1999 -0.02082 0.06126 -0.34 0.734
## Year2000 0.01169 0.06780 0.17 0.863
## Year2001 -0.07920 0.07234 -1.09 0.274
## Year2002 -0.11133 0.06473 -1.72 0.086 .
## Year2003 0.00216 0.06089 0.04 0.972
## Year2004 -0.05583 0.06332 -0.88 0.378
## Year2005 -0.11202 0.06665 -1.68 0.093 .
## Year2006 -0.06418 0.05771 -1.11 0.266
```

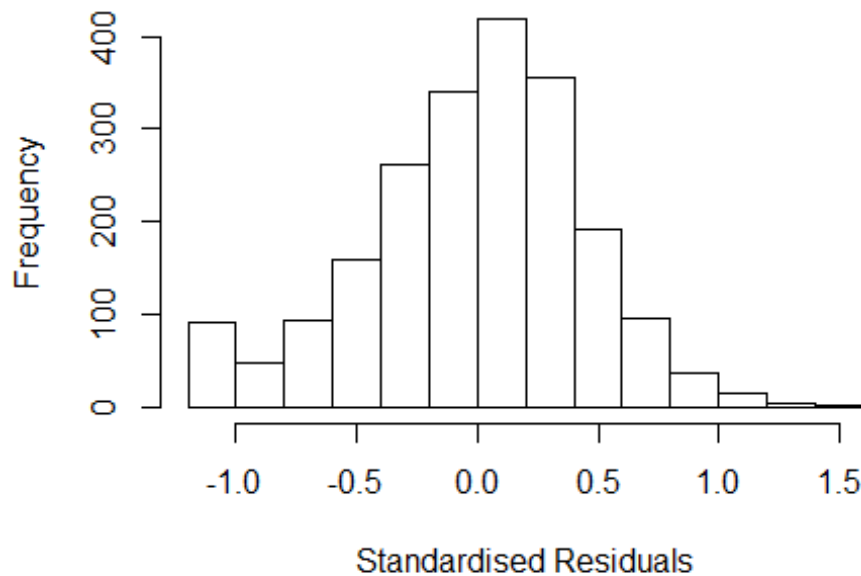


```

## Year2007          -0.16504    0.06615   -2.50    0.013 *
## Year2008          -0.04716    0.06086   -0.77    0.439
## Year2009          -0.13796    0.06680   -2.07    0.039 *
## Year2010          -0.11270    0.06241   -1.81    0.071 .
## Year2011          -0.07697    0.06334   -1.22    0.224
## Year2012          -0.06657    0.06194   -1.07    0.283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0116, Adjusted R-squared:  0.00357
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1941 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.201  0.859  0.950  0.893  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1      1.017
## Year              1.035 16      1.001

```

## Residuals from last author



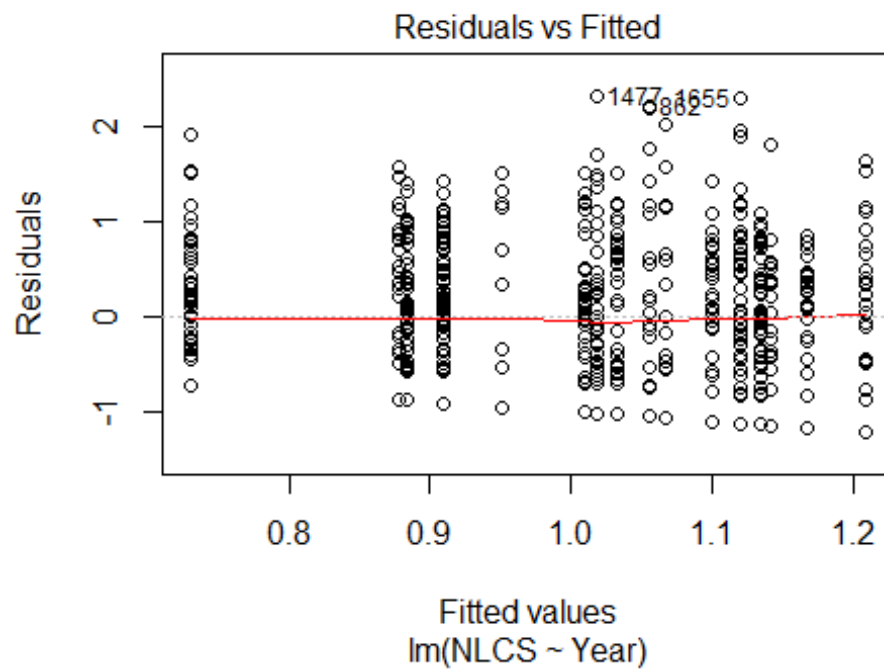
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1955 -0.2849 0.0258 0.2784 1.4628
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18456 0.04934 24.01 <2e-16 ***
## LastAuthorFemale1 -0.01988 0.03082 -0.65 0.519
## Year1997 -0.05766 0.06517 -0.88 0.376
## Year1998 -0.03405 0.06445 -0.53 0.597
## Year1999 -0.02125 0.06134 -0.35 0.729
## Year2000 0.01092 0.06793 0.16 0.872
## Year2001 -0.08115 0.07260 -1.12 0.264
## Year2002 -0.11180 0.06482 -1.72 0.085 .
## Year2003 0.00201 0.06096 0.03 0.974
## Year2004 -0.05758 0.06351 -0.91 0.365
## Year2005 -0.11232 0.06669 -1.68 0.092 .
## Year2006 -0.06430 0.05779 -1.11 0.266
```

```

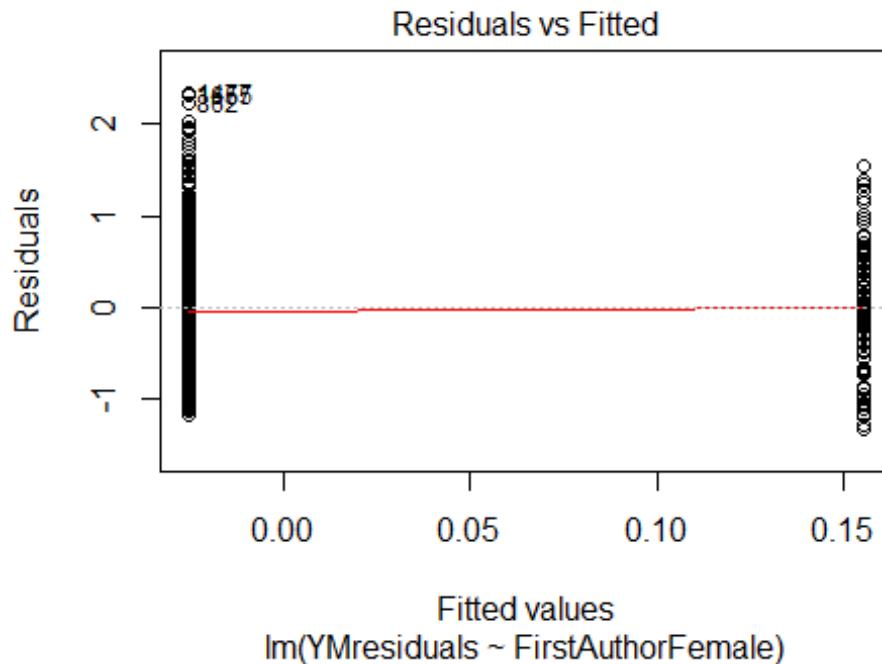
## Year2007          -0.16474      0.06619    -2.49      0.013 *
## Year2008          -0.04770      0.06084    -0.78      0.433
## Year2009          -0.13759      0.06703    -2.05      0.040 *
## Year2010          -0.11261      0.06250    -1.80      0.072 .
## Year2011          -0.07831      0.06347    -1.23      0.217
## Year2012          -0.06738      0.06201    -1.09      0.277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.0117, Adjusted R-squared:  0.00373
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 174 weights are ~= 1. The remaining 1941 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.202  0.859  0.950  0.893  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2115"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1909"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 112 130 143 140 195 140 179 118 122 126 137 130 110 127
## 2011 2012
## 135 147
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 18 29 28 25 31 36 65 44 54 37 48 47 34 50
## 2011 2012

```

```
## 49 50
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 15 26 25 22 27 34 60 41 45 31 37 38 29 48
## 2011 2012
## 46 46
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```

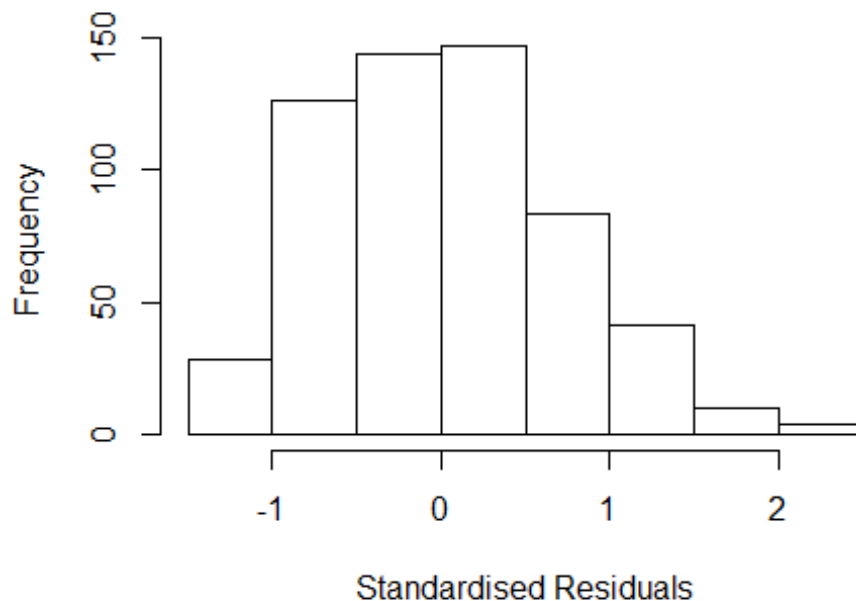


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.18, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 2.229989887763"
## [1] "Male first author team size 2018 geometric mean: 2.35438873453321"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 810, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.91706972723258"
## [1] "Male last author team size 2018 geometric mean: 2.21250994087362"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.332 1          1.154
## LastAuthorFemale  1.368 1          1.170
## UniqueAuthors     2.105 4          1.097
## Year              1.934 16         1.021
```

## Residuals from first and last author and team size



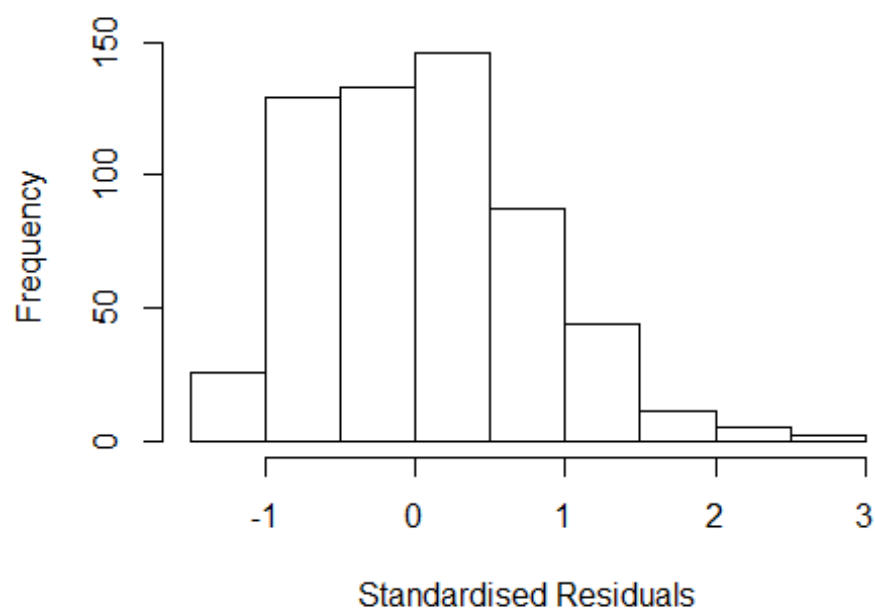
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4516 -0.5126 -0.0185 0.4596 2.4462
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5245 0.2888 1.82 0.0699 .
## FirstAuthorFemale1 0.2042 0.1036 1.97 0.0492 *
## LastAuthorFemale1 -0.0792 0.1044 -0.76 0.4484
## UniqueAuthors2 0.5459 0.0840 6.50 1.8e-10 ***
## UniqueAuthors3 0.5592 0.0884 6.33 5.1e-10 ***
## UniqueAuthors4 0.4109 0.1413 2.91 0.0038 **
## UniqueAuthors5 0.6503 0.0926 7.02 6.3e-12 ***
## Year1997 0.1009 0.4114 0.25 0.8063
## Year1998 0.1581 0.3345 0.47 0.6367
## Year1999 0.3679 0.3439 1.07 0.2851
```

```

## Year2000      0.2783      0.3242      0.86      0.3909
## Year2001      0.1911      0.3803      0.50      0.6155
## Year2002      0.1170      0.3303      0.35      0.7233
## Year2003     -0.1034      0.3026     -0.34      0.7328
## Year2004     -0.0375      0.3073     -0.12      0.9030
## Year2005      0.0463      0.3154      0.15      0.8834
## Year2006      0.2731      0.3184      0.86      0.3914
## Year2007      0.2202      0.3208      0.69      0.4926
## Year2008      0.2672      0.3056      0.87      0.3822
## Year2009      0.2726      0.3166      0.86      0.3895
## Year2010      0.0736      0.3002      0.25      0.8064
## Year2011     -0.0579      0.3059     -0.19      0.8498
## Year2012     -0.0119      0.3048     -0.04      0.9688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.67
## Multiple R-squared:  0.181, Adjusted R-squared:  0.149
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 539 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.155  0.875  0.941  0.902  0.979  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.198 1      1.094
## LastAuthorFemale  1.152 1      1.073
## Year              1.107 16      1.003

```

## Residuals from first and last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1477  3042826336 3.330 2004    1909      1    2.546
## 1655 15544382269 3.417 2005    1909      1    2.510
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.402 -0.586  0.013  0.515  2.546
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7555     0.3422   2.21  0.0277 *
## FirstAuthorFemale1 0.3075     0.1070   2.87  0.0042 **
## LastAuthorFemale1 -0.0658     0.1108  -0.59  0.5526
## Year1997         0.0800     0.4486   0.18  0.8586
## Year1998         0.0947     0.3895   0.24  0.8080
## Year1999         0.4252     0.3885   1.09  0.2742
## Year2000         0.3521     0.3612   0.97  0.3301
## Year2001         0.1194     0.4091   0.29  0.7704
## Year2002         0.1356     0.3800   0.36  0.7214
## Year2003        -0.1310     0.3528  -0.37  0.7106
## Year2004         0.0284     0.3605   0.08  0.9372
```

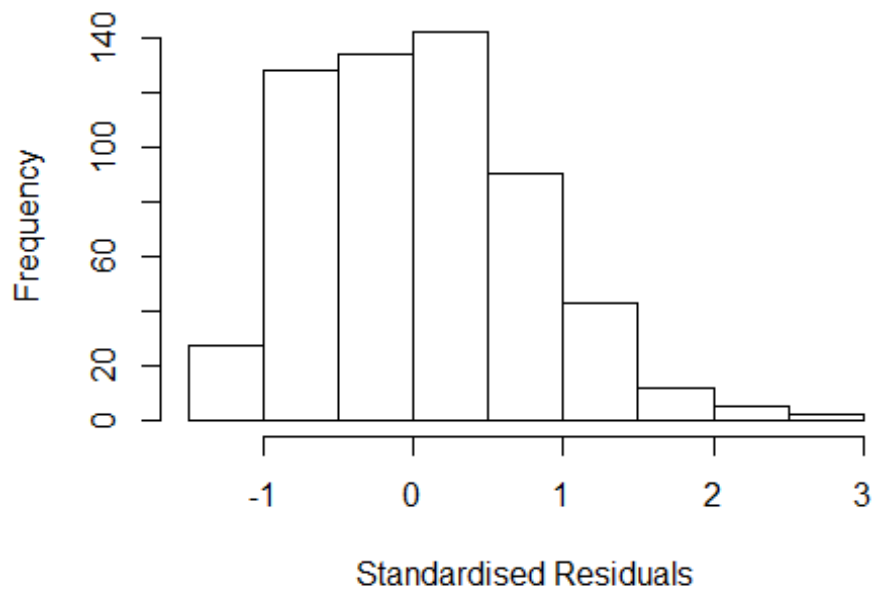


```

## Year2005      0.1518      0.3667      0.41      0.6790
## Year2006      0.3870      0.3693      1.05      0.2951
## Year2007      0.2340      0.3686      0.63      0.5259
## Year2008      0.3765      0.3565      1.06      0.2913
## Year2009      0.4050      0.3647      1.11      0.2672
## Year2010      0.0894      0.3545      0.25      0.8010
## Year2011      0.0587      0.3572      0.16      0.8695
## Year2012      0.1008      0.3582      0.28      0.7785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.731
## Multiple R-squared:  0.0604, Adjusted R-squared:  0.0304
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 541 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.200  0.874   0.937   0.906   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.083 1      1.041
## Year      1.083 16      1.002

```

## Residuals from first author

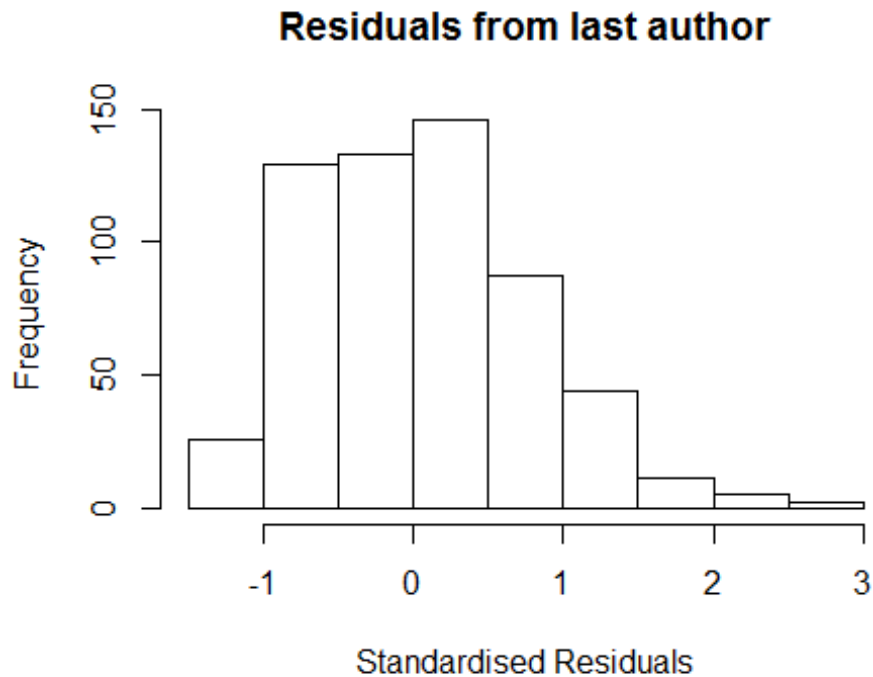


```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1477  3042826336 3.330 2004    1909      1    2.546
## 1655 15544382269 3.417 2005    1909      1    2.510
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4432 -0.5896  0.0142  0.5166  2.5520
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7524    0.3462   2.17  0.0302 *
## FirstAuthorFemale1 0.2839    0.1026   2.77  0.0058 **
## Year1997        0.0823    0.4516   0.18  0.8555
## Year1998        0.0952    0.3930   0.24  0.8086
## Year1999        0.4201    0.3922   1.07  0.2845
## Year2000        0.3530    0.3655   0.97  0.3345
## Year2001        0.1201    0.4123   0.29  0.7709
## Year2002        0.1350    0.3834   0.35  0.7250
## Year2003       -0.1349    0.3568  -0.38  0.7055
## Year2004        0.0256    0.3647   0.07  0.9441
## Year2005        0.1502    0.3708   0.40  0.6856
```

```

## Year2006          0.3855      0.3736      1.03      0.3027
## Year2007          0.2342      0.3727      0.63      0.5301
## Year2008          0.3774      0.3607      1.05      0.2959
## Year2009          0.4069      0.3688      1.10      0.2705
## Year2010          0.0887      0.3586      0.25      0.8048
## Year2011          0.0566      0.3611      0.16      0.8756
## Year2012          0.1014      0.3625      0.28      0.7798
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.73
## Multiple R-squared:  0.06,   Adjusted R-squared:  0.0318
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 536 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.196  0.874   0.936   0.905   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.031 1      1.015
## Year      1.031 16      1.001

```



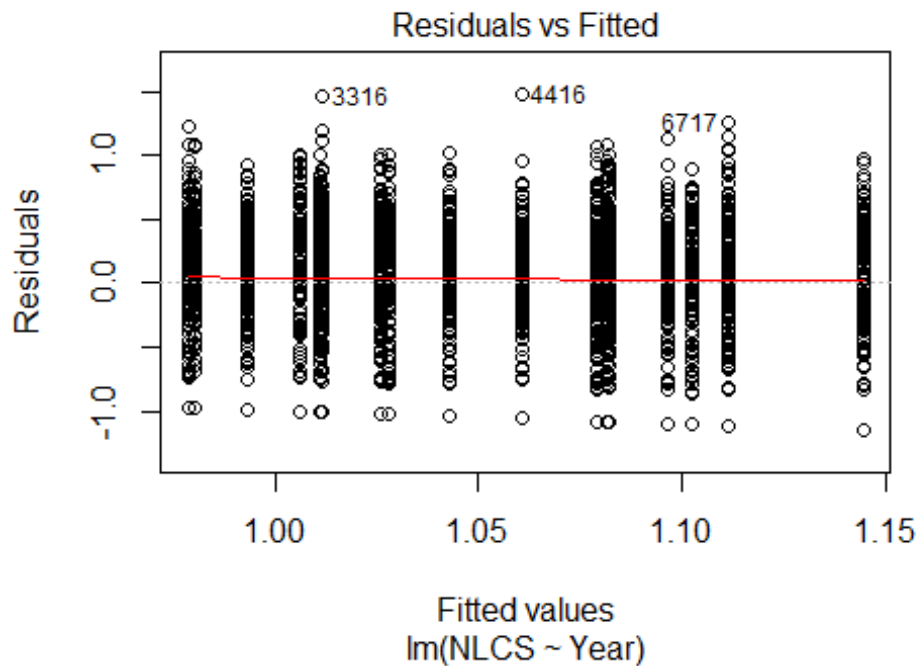
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1477 3042826336 3.330 2004    1909      1    2.546
## 1655 15544382269 3.417 2005    1909      1    2.510
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2726 -0.5831  0.0106  0.5415  2.5214
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7722    0.3472   2.22  0.027 *
## LastAuthorFemale1 0.0676    0.1010   0.67  0.504
## Year1997        0.0666    0.4512   0.15  0.883
## Year1998        0.0881    0.3938   0.22  0.823
## Year1999        0.3923    0.3924   1.00  0.318
## Year2000        0.3409    0.3676   0.93  0.354
## Year2001        0.1308    0.4096   0.32  0.750
## Year2002        0.1864    0.3785   0.49  0.623
## Year2003       -0.1197    0.3564  -0.34  0.737
## Year2004        0.0364    0.3659   0.10  0.921
## Year2005        0.1846    0.3719   0.50  0.620
```

```

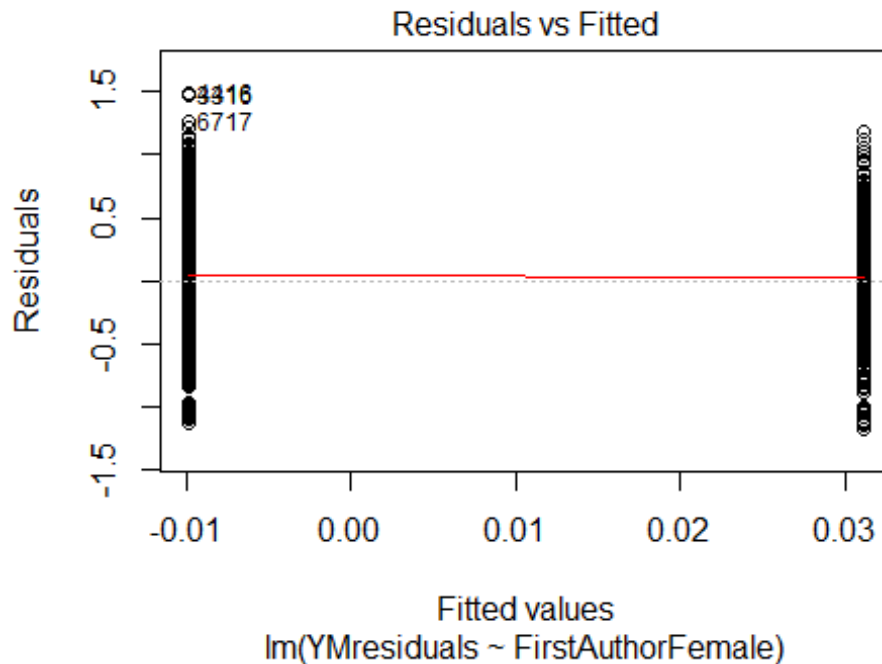
## Year2006          0.4052      0.3750      1.08      0.280
## Year2007          0.2388      0.3722      0.64      0.521
## Year2008          0.3916      0.3617      1.08      0.279
## Year2009          0.4328      0.3681      1.18      0.240
## Year2010          0.0998      0.3587      0.28      0.781
## Year2011          0.0485      0.3603      0.13      0.893
## Year2012          0.1183      0.3622      0.33      0.744
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.738
## Multiple R-squared:  0.0452, Adjusted R-squared:  0.0165
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 545 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.219  0.872  0.934  0.907  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 583"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1910"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   298  324  315  369  324  356  367  294  260  240  270  313  288  278  288
## 2011 2012
##   289  268
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   121  121  121  177  133  130  184  150  151  157  166  198  184  177  174

```

```
## 2011 2012
## 183 158
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 114 113 111 165 118 114 163 138 144 147 151 171 162 161 154
## 2011 2012
## 161 148
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 6e-04
```

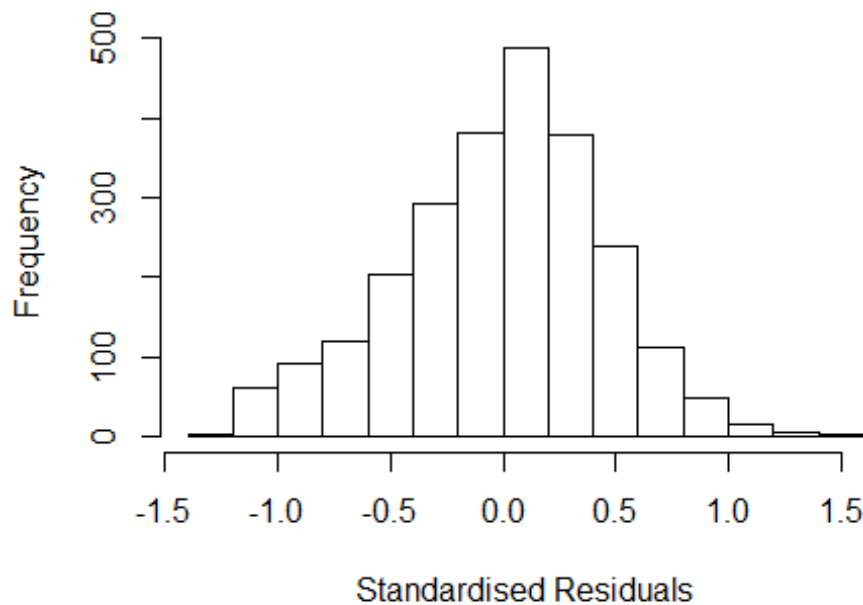


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 17, df = 1, p-value = 3e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.22806973004818"
## [1] "Male first author team size 2018 geometric mean: 2.70884422381731"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.15771550428087"
## [1] "Male last author team size 2018 geometric mean: 2.78766552525302"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.067 1          1.033
## LastAuthorFemale  1.052 1          1.026
## UniqueAuthors    1.143 4          1.017
## Year              1.181 16         1.005
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2353 -0.2922 0.0278 0.2934 1.5414
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90977 0.05031 18.08 <2e-16 ***
## FirstAuthorFemale1 0.00926 0.02090 0.44 0.658
## LastAuthorFemale1 -0.03258 0.02566 -1.27 0.204
## UniqueAuthors2 0.27467 0.02918 9.41 <2e-16 ***
## UniqueAuthors3 0.31632 0.02991 10.58 <2e-16 ***
## UniqueAuthors4 0.32965 0.03298 10.00 <2e-16 ***
## UniqueAuthors5 0.43498 0.03156 13.78 <2e-16 ***
## Year1997 -0.02713 0.07081 -0.38 0.702
## Year1998 -0.14431 0.07015 -2.06 0.040 *
## Year1999 -0.07168 0.06213 -1.15 0.249
```

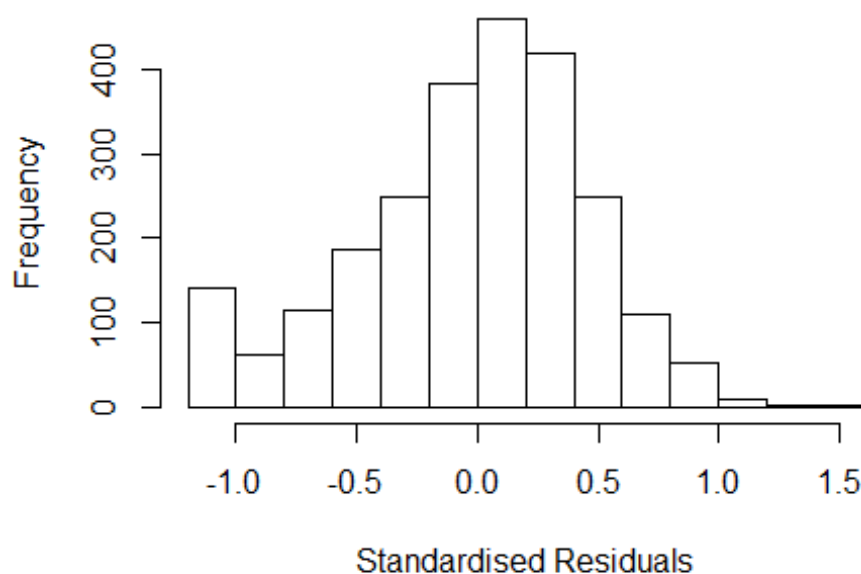


```

## Year2000      -0.00338    0.06134   -0.06    0.956
## Year2001      -0.13334    0.07280   -1.83    0.067 .
## Year2002      -0.08326    0.06222   -1.34    0.181
## Year2003      -0.07770    0.06438   -1.21    0.228
## Year2004      -0.12426    0.05860   -2.12    0.034 *
## Year2005      -0.12481    0.06170   -2.02    0.043 *
## Year2006      -0.08652    0.05704   -1.52    0.129
## Year2007      -0.05640    0.05733   -0.98    0.325
## Year2008      -0.07954    0.05494   -1.45    0.148
## Year2009      -0.12481    0.05639   -2.21    0.027 *
## Year2010      -0.06661    0.06044   -1.10    0.271
## Year2011      -0.08715    0.05766   -1.51    0.131
## Year2012      -0.01152    0.05980   -0.19    0.847
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.43
## Multiple R-squared:  0.11, Adjusted R-squared:  0.102
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 214 weights are ~= 1. The remaining 2221 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.172  0.862  0.949   0.896   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.056 1      1.028
## LastAuthorFemale  1.049 1      1.024
## Year              1.062 16      1.002

```

## Residuals from first and last author



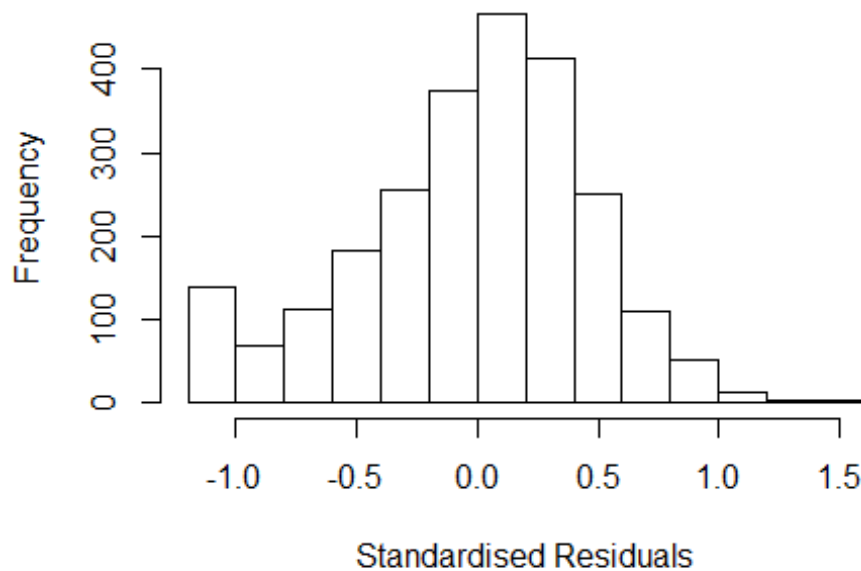
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1659 -0.3108 0.0405 0.2993 1.4274
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10552 0.04794 23.06 <2e-16 ***
## FirstAuthorFemale1 0.05372 0.02155 2.49 0.013 *
## LastAuthorFemale1 -0.04604 0.02688 -1.71 0.087 .
## Year1997 -0.03716 0.07161 -0.52 0.604
## Year1998 -0.13409 0.07563 -1.77 0.076 .
## Year1999 -0.07065 0.06575 -1.07 0.283
## Year2000 0.01933 0.06409 0.30 0.763
## Year2001 -0.10004 0.08002 -1.25 0.211
## Year2002 -0.04699 0.06344 -0.74 0.459
## Year2003 -0.05996 0.06559 -0.91 0.361
## Year2004 -0.09609 0.06003 -1.60 0.110
## Year2005 -0.11779 0.06234 -1.89 0.059 .
```

```

## Year2006          -0.04706      0.05847    -0.80      0.421
## Year2007          -0.01451      0.05983    -0.24      0.808
## Year2008          -0.01801      0.05705    -0.32      0.752
## Year2009          -0.06796      0.05947    -1.14      0.253
## Year2010          -0.01764      0.06199    -0.28      0.776
## Year2011          -0.00739      0.05962    -0.12      0.901
## Year2012           0.05267      0.06172      0.85      0.394
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.013, Adjusted R-squared:  0.00565
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 200 weights are ~= 1. The remaining 2235 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.264  0.857  0.948  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.035 1      1.017
## Year              1.035 16      1.001

```

## Residuals from first author



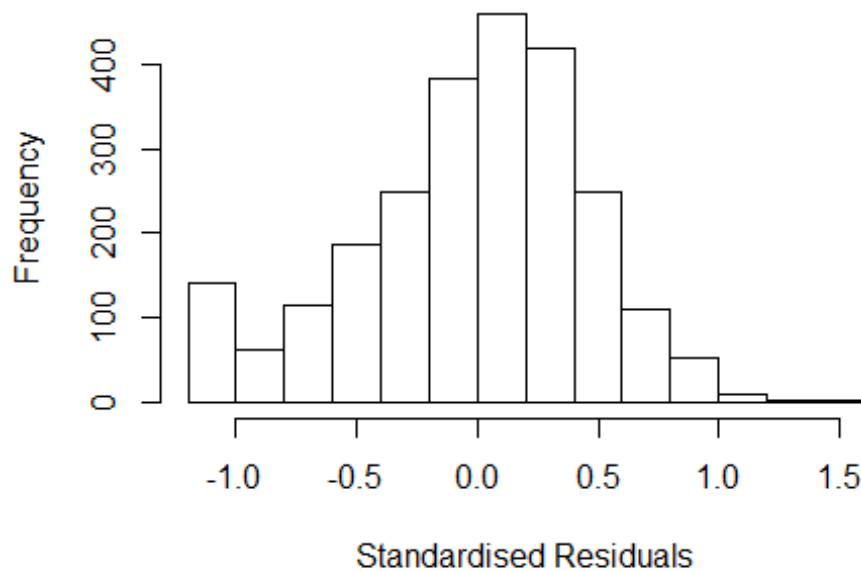
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1974 -0.3081 0.0402 0.3026 1.4326
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1014 0.0478 23.04 <2e-16 ***
## FirstAuthorFemale1 0.0442 0.0216 2.05 0.040 *
## Year1997 -0.0411 0.0719 -0.57 0.567
## Year1998 -0.1337 0.0755 -1.77 0.077 .
## Year1999 -0.0714 0.0658 -1.09 0.278
## Year2000 0.0175 0.0643 0.27 0.785
## Year2001 -0.1028 0.0800 -1.28 0.199
## Year2002 -0.0479 0.0635 -0.75 0.451
## Year2003 -0.0610 0.0656 -0.93 0.353
## Year2004 -0.0965 0.0600 -1.61 0.108
## Year2005 -0.1180 0.0624 -1.89 0.059 .
## Year2006 -0.0477 0.0584 -0.82 0.415
```

```

## Year2007          -0.0188      0.0597   -0.31    0.753
## Year2008          -0.0176      0.0571   -0.31    0.757
## Year2009          -0.0704      0.0593   -1.19    0.236
## Year2010          -0.0166      0.0620   -0.27    0.789
## Year2011          -0.0077      0.0596   -0.13    0.897
## Year2012           0.0518      0.0617    0.84    0.401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0119, Adjusted R-squared:  0.00493
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 189 weights are ~= 1. The remaining 2246 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.260  0.858  0.949   0.891  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1          1.014
## Year              1.028 16          1.001

```

## Residuals from last author



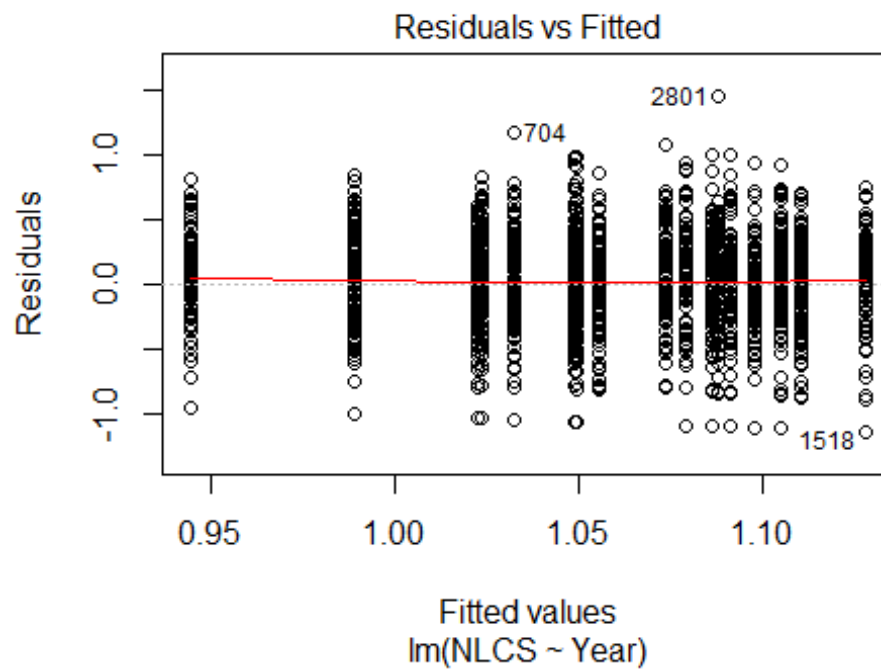
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1751 -0.3126 0.0402 0.3020 1.4158
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11024 0.04819 23.04 <2e-16 ***
## LastAuthorFemale1 -0.03079 0.02645 -1.16 0.245
## Year1997 -0.03330 0.07189 -0.46 0.643
## Year1998 -0.13190 0.07585 -1.74 0.082 .
## Year1999 -0.06721 0.06598 -1.02 0.308
## Year2000 0.02320 0.06399 0.36 0.717
## Year2001 -0.09561 0.08008 -1.19 0.233
## Year2002 -0.03954 0.06341 -0.62 0.533
## Year2003 -0.05304 0.06580 -0.81 0.420
## Year2004 -0.09151 0.06014 -1.52 0.128
## Year2005 -0.10994 0.06249 -1.76 0.079 .
## Year2006 -0.03786 0.05825 -0.65 0.516
```

```

## Year2007          -0.00913      0.05995    -0.15      0.879
## Year2008          -0.01099      0.05706    -0.19      0.847
## Year2009          -0.05943      0.05944    -1.00      0.318
## Year2010          -0.00838      0.06201    -0.14      0.893
## Year2011           0.00279      0.05939      0.05      0.963
## Year2012           0.06491      0.06135      1.06      0.290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00387
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 193 weights are ~= 1. The remaining 2242 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.272  0.858  0.947  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2435"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1911"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 188 204 193 198 237 176 170 156 177 130 153 185 187 173 182
## 2011 2012
## 193 174
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 113 118 124 126 78 107 102 120 94 96 123 132 118 118
## 2011 2012

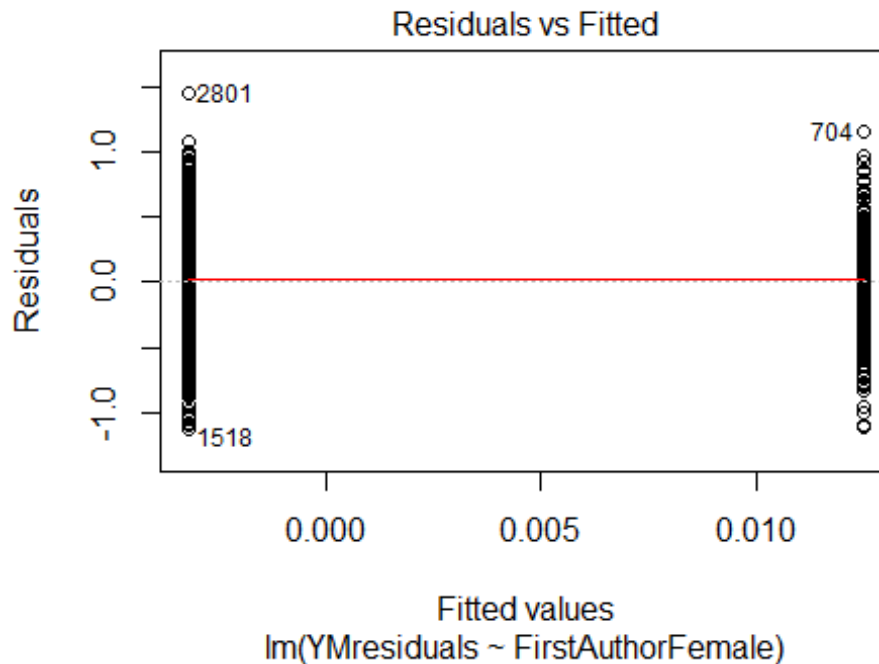
```

```
## 120 111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 96 106 109 121 116 63 91 94 113 86 88 111 120 112 111
## 2011 2012
## 111 98
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.04
```



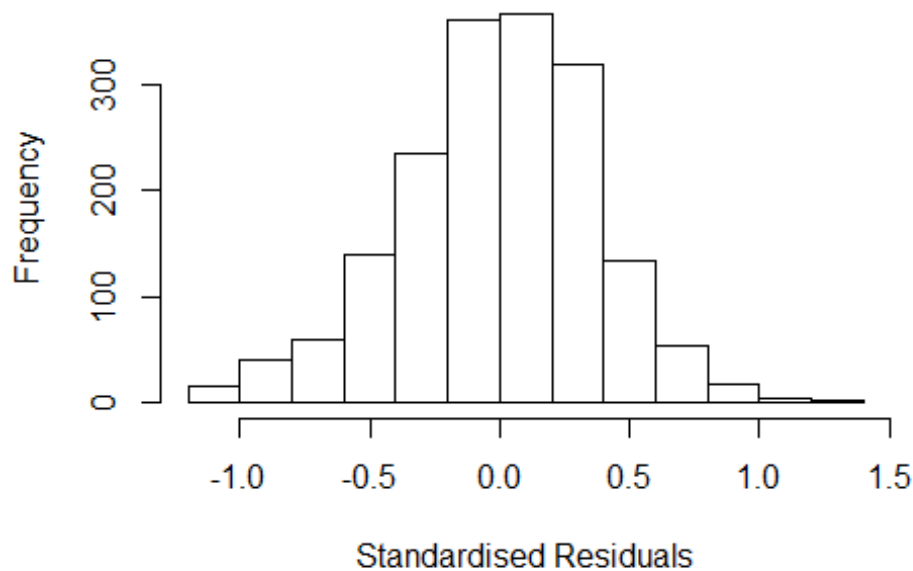
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
```





```
## [1] "Female first author team size 2018 geometric mean: 2.9213804990243"
## [1] "Male first author team size 2018 geometric mean: 2.38480288449846"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.17485428845557"
## [1] "Male last author team size 2018 geometric mean: 2.54128738379041"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 590, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.196 1      1.094
## LastAuthorFemale  1.208 1      1.099
## UniqueAuthors    1.254 4      1.029
## Year              1.297 16     1.008
```

## Residuals from first and last author and team size



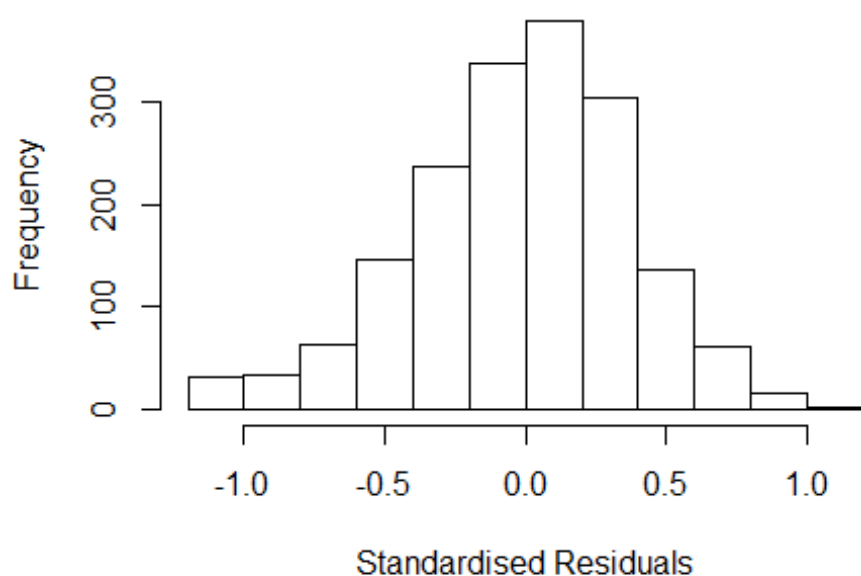
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1675 -0.2422 0.0143 0.2437 1.2873
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.948536 0.043076 22.02 < 2e-16 ***
## FirstAuthorFemale1 0.009517 0.022908 0.42 0.68
## LastAuthorFemale1 -0.036890 0.025381 -1.45 0.15
## UniqueAuthors2 0.163600 0.023734 6.89 7.6e-12 ***
## UniqueAuthors3 0.188981 0.026105 7.24 6.8e-13 ***
## UniqueAuthors4 0.243605 0.031982 7.62 4.3e-14 ***
## UniqueAuthors5 0.277803 0.032673 8.50 < 2e-16 ***
## Year1997 -0.000662 0.058481 -0.01 0.99
## Year1998 -0.009458 0.057071 -0.17 0.87
## Year1999 -0.029637 0.056544 -0.52 0.60
```

```

## Year2000      -0.039057    0.052111   -0.75    0.45
## Year2001      0.049804    0.063595    0.78    0.43
## Year2002      0.022274    0.060326    0.37    0.71
## Year2003     -0.067752    0.058445   -1.16    0.25
## Year2004     -0.019001    0.050582   -0.38    0.71
## Year2005      0.046976    0.053368    0.88    0.38
## Year2006      0.000749    0.055593    0.01    0.99
## Year2007      0.042993    0.051856    0.83    0.41
## Year2008      0.057357    0.053977    1.06    0.29
## Year2009     -0.013767    0.052404   -0.26    0.79
## Year2010     -0.006063    0.055520   -0.11    0.91
## Year2011     -0.023421    0.055496   -0.42    0.67
## Year2012     -0.000484    0.056692   -0.01    0.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.361
## Multiple R-squared:  0.078, Adjusted R-squared:  0.0662
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1605 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.177  0.871  0.952   0.902   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.187 1      1.089
## LastAuthorFemale  1.180 1      1.086
## Year              1.084 16      1.003

```

## Residuals from first and last author



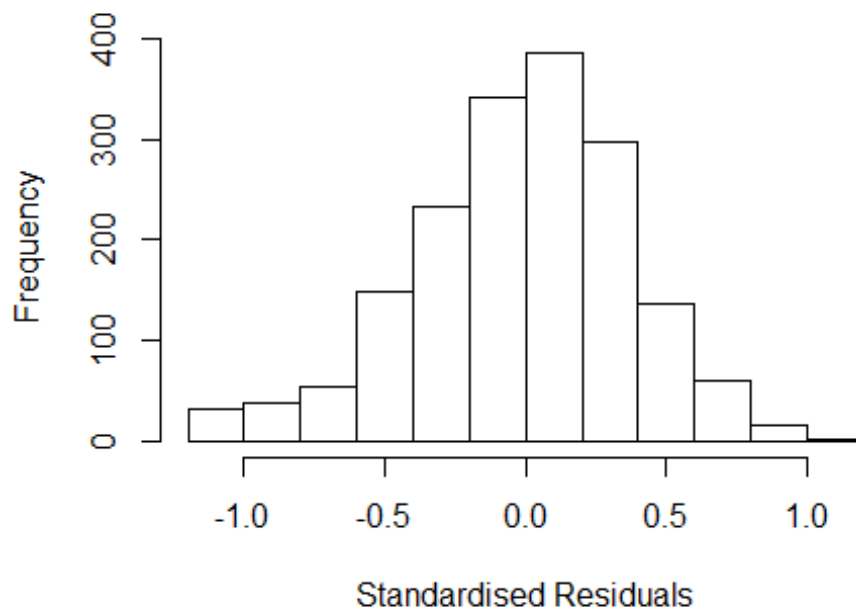
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1274 -0.2489  0.0169  0.2500  1.1483
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.04243    0.04475   23.30  <2e-16 ***
## FirstAuthorFemale1  0.03396    0.02320    1.46    0.14
## LastAuthorFemale1 -0.04128    0.02593   -1.59    0.11
## Year1997          0.00917    0.06108    0.15    0.88
## Year1998          0.01557    0.05954    0.26    0.79
## Year1999         -0.02912    0.06126   -0.48    0.63
## Year2000          0.00153    0.05519    0.03    0.98
## Year2001          0.09805    0.06679    1.47    0.14
## Year2002          0.05373    0.06147    0.87    0.38
## Year2003         -0.05833    0.06039   -0.97    0.33
## Year2004         -0.00884    0.05449   -0.16    0.87
## Year2005          0.05266    0.05578    0.94    0.35
```

```

## Year2006          0.02593    0.05742    0.45    0.65
## Year2007          0.08416    0.05464    1.54    0.12
## Year2008          0.08495    0.05742    1.48    0.14
## Year2009          0.02106    0.05596    0.38    0.71
## Year2010          0.05505    0.05897    0.93    0.35
## Year2011          0.03030    0.05787    0.52    0.60
## Year2012          0.05972    0.06036    0.99    0.32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0133, Adjusted R-squared:  0.00303
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1609 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.315  0.869  0.952  0.900  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1      1.021
## Year              1.043 16      1.001

```

## Residuals from first author



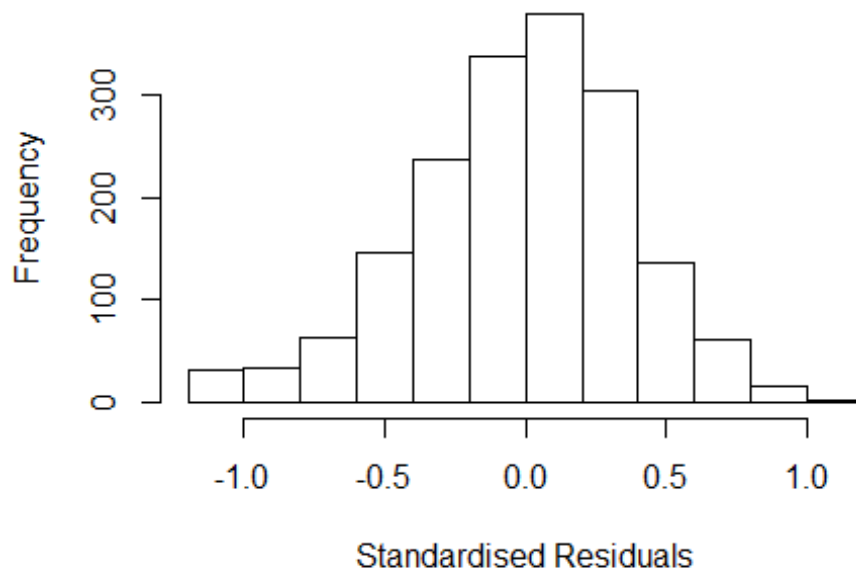
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1477 -0.2454 0.0156 0.2503 1.1213
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03404 0.04411 23.44 <2e-16 ***
## FirstAuthorFemale1 0.02201 0.02184 1.01 0.31
## Year1997 0.01100 0.06111 0.18 0.86
## Year1998 0.02168 0.05917 0.37 0.71
## Year1999 -0.02380 0.06091 -0.39 0.70
## Year2000 0.00216 0.05513 0.04 0.97
## Year2001 0.10373 0.06642 1.56 0.12
## Year2002 0.05892 0.06120 0.96 0.34
## Year2003 -0.05397 0.06044 -0.89 0.37
## Year2004 -0.00502 0.05436 -0.09 0.93
## Year2005 0.05853 0.05541 1.06 0.29
## Year2006 0.03107 0.05705 0.54 0.59
```

```

## Year2007          0.08545    0.05468    1.56    0.12
## Year2008          0.09160    0.05691    1.61    0.11
## Year2009          0.02544    0.05562    0.46    0.65
## Year2010          0.05959    0.05897    1.01    0.31
## Year2011          0.03528    0.05771    0.61    0.54
## Year2012          0.06461    0.06009    1.08    0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00227
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 133 weights are ~= 1. The remaining 1613 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.315  0.868  0.952  0.900  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.037 1          1.018
## Year            1.037 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1338 -0.2490 0.0162 0.2475 1.1643
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04581 0.04462 23.44 <2e-16 ***
## LastAuthorFemale1 -0.02939 0.02423 -1.21 0.23
## Year1997 0.00838 0.06084 0.14 0.89
## Year1998 0.01831 0.05939 0.31 0.76
## Year1999 -0.02888 0.06131 -0.47 0.64
## Year2000 0.00133 0.05519 0.02 0.98
## Year2001 0.10035 0.06643 1.51 0.13
## Year2002 0.05366 0.06145 0.87 0.38
## Year2003 -0.05634 0.06041 -0.93 0.35
## Year2004 -0.00663 0.05434 -0.12 0.90
## Year2005 0.05803 0.05545 1.05 0.30
## Year2006 0.02740 0.05733 0.48 0.63
```

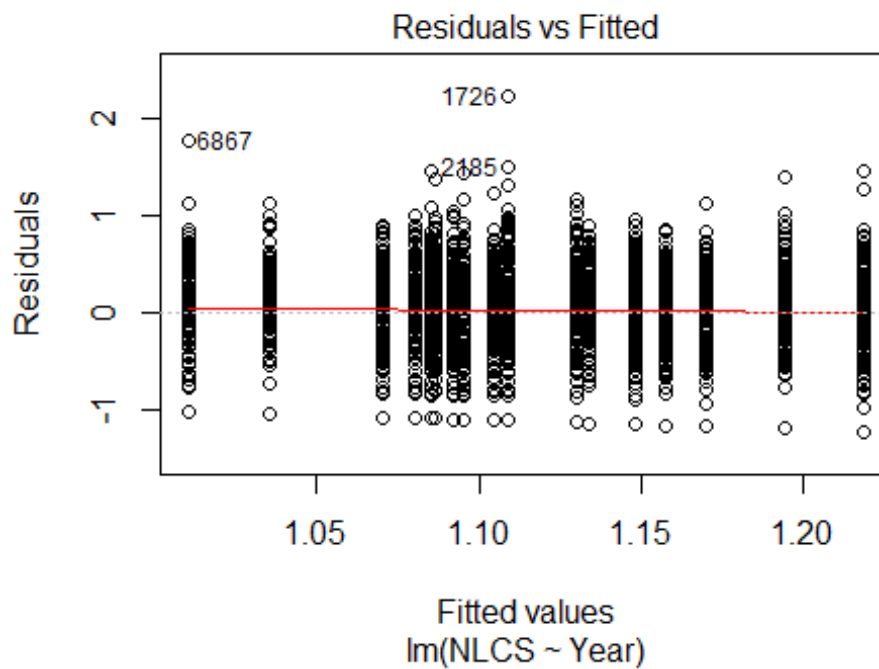


```

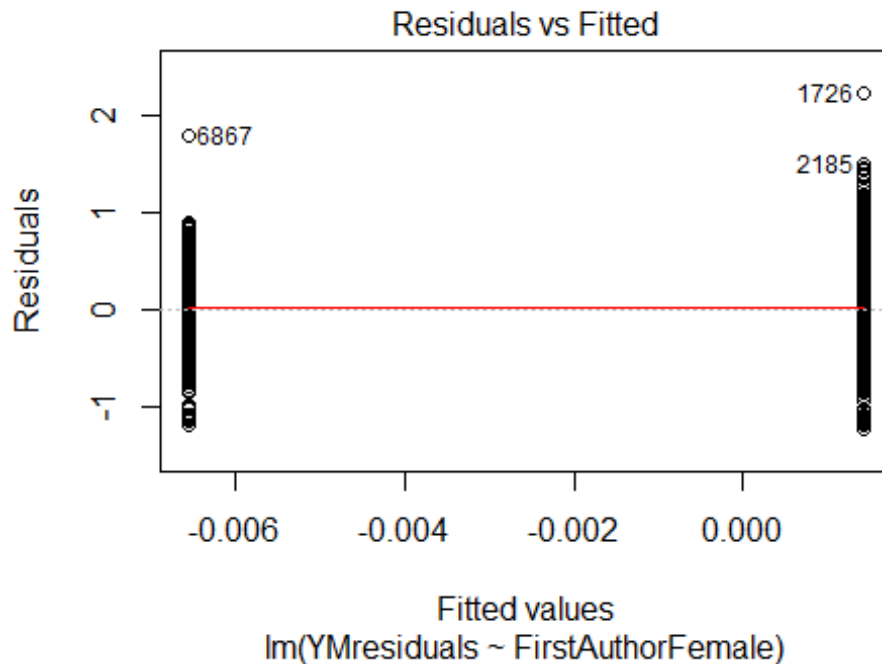
## Year2007      0.08523      0.05466      1.56      0.12
## Year2008      0.08802      0.05718      1.54      0.12
## Year2009      0.02255      0.05598      0.40      0.69
## Year2010      0.05757      0.05878      0.98      0.33
## Year2011      0.03473      0.05761      0.60      0.55
## Year2012      0.06275      0.06025      1.04      0.30
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0122, Adjusted R-squared:  0.0025
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1608 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.301  0.872  0.952  0.900  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.73e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1746"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1912"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  507  552  535  521  477  520  526  492  465  456  505  459  452  493  466
## 2011 2012
##  548  458
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  137  166  164  161  148  146  164  161  143  163  155  156  163  185  174
## 2011 2012

```

```
## 212 190
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 118 146 146 148 137 135 144 146 135 138 142 131 140 166 158
## 2011 2012
## 189 173
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 51, df = 16, p-value = 2e-05
```

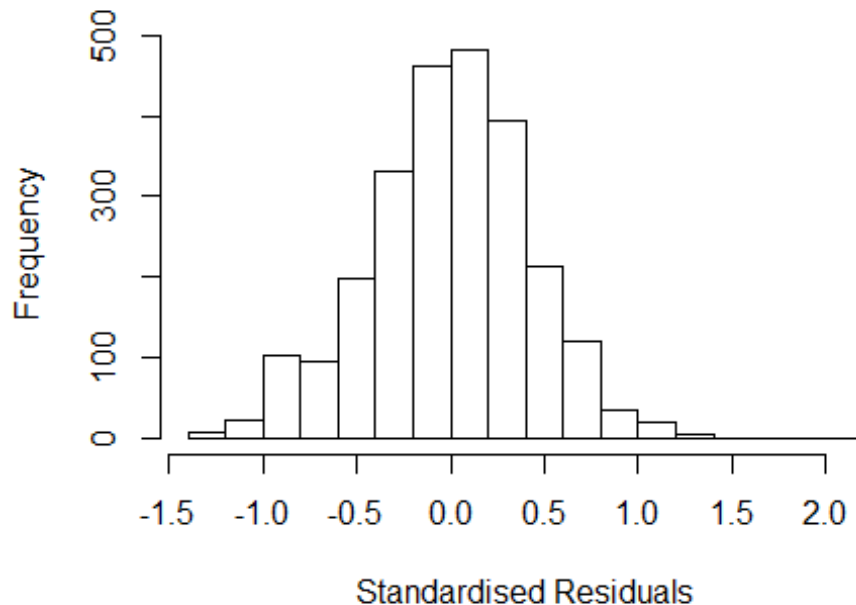


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.9, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 2.98915818919678"
## [1] "Male first author team size 2018 geometric mean: 2.60791725026912"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.60797120642035"
## [1] "Male last author team size 2018 geometric mean: 2.68310300052557"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 990, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1 1.021
## LastAuthorFemale 1.036 1 1.018
## UniqueAuthors 1.179 4 1.021
## Year 1.200 16 1.006
```

## Residuals from first and last author and team size



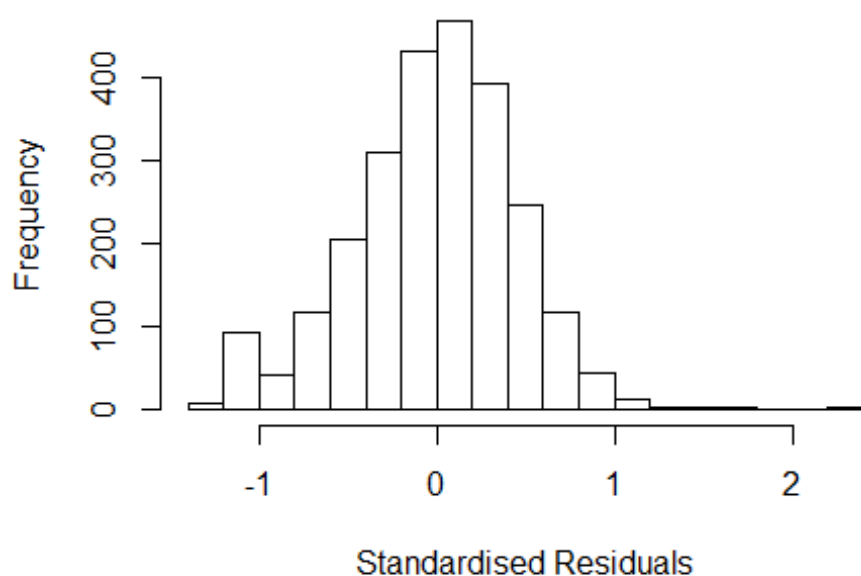
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29414 -0.26701  0.00911  0.27543  2.11788
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91328    0.04324   21.12  <2e-16 ***
## FirstAuthorFemale1 -0.05334    0.02182   -2.44   0.015 *
## LastAuthorFemale1 -0.01047    0.02739   -0.38   0.702
## UniqueAuthors2    0.26422    0.02517   10.50  <2e-16 ***
## UniqueAuthors3    0.28652    0.02747   10.43  <2e-16 ***
## UniqueAuthors4    0.28890    0.03274    8.83  <2e-16 ***
## UniqueAuthors5    0.38086    0.03603   10.57  <2e-16 ***
## Year1997          0.04258    0.05840    0.73   0.466
## Year1998          0.01431    0.06194    0.23   0.817
## Year1999         -0.01410    0.05642   -0.25   0.803
```

```

## Year2000      0.11786      0.05352      2.20      0.028 *
## Year2001      0.01087      0.05633      0.19      0.847
## Year2002     -0.04657      0.05446     -0.86      0.393
## Year2003     -0.08178      0.05903     -1.39      0.166
## Year2004      0.05997      0.05439      1.10      0.270
## Year2005      0.05577      0.05511      1.01      0.312
## Year2006     -0.04355      0.05638     -0.77      0.440
## Year2007     -0.02136      0.05646     -0.38      0.705
## Year2008     -0.00169      0.05368     -0.03      0.975
## Year2009      0.05588      0.05193      1.08      0.282
## Year2010     -0.02254      0.04938     -0.46      0.648
## Year2011     -0.01207      0.05034     -0.24      0.811
## Year2012      0.05887      0.05090      1.16      0.248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0981, Adjusted R-squared:  0.09
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 281 is an outlier with |weight| = 0 ( < 4e-05);
## 211 weights are ~= 1. The remaining 2280 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0268 0.8650 0.9520 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max      maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1      1.008
## LastAuthorFemale  1.035 1      1.017
## Year              1.045 16      1.001

```

## Residuals from first and last author



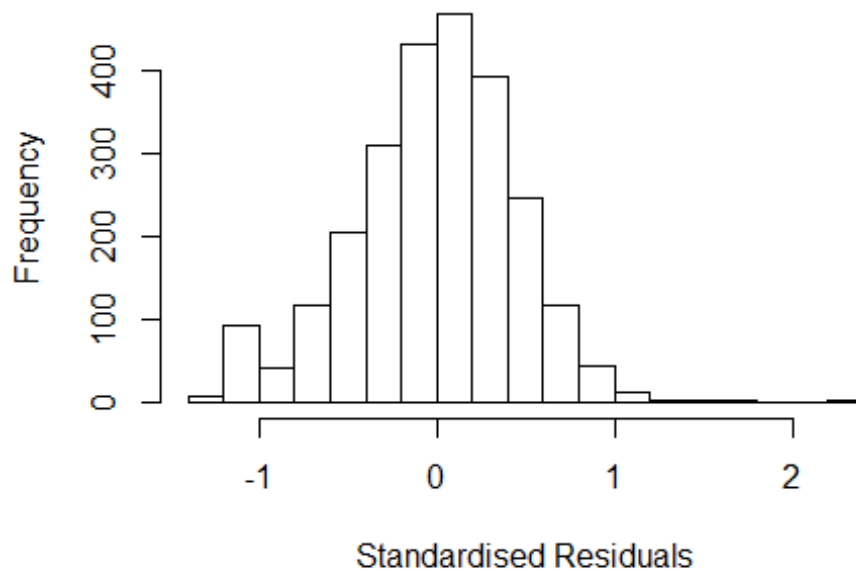
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2264 -0.2829 0.0173 0.2860 2.2241
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08e+00 4.35e-02 24.84 <2e-16 ***
## FirstAuthorFemale1 -6.67e-03 2.21e-02 -0.30 0.763
## LastAuthorFemale1 -6.64e-05 2.86e-02 0.00 0.998
## Year1997 1.85e-02 6.00e-02 0.31 0.758
## Year1998 2.73e-02 6.47e-02 0.42 0.673
## Year1999 -1.09e-02 5.95e-02 -0.18 0.854
## Year2000 1.46e-01 5.66e-02 2.57 0.010 *
## Year2001 3.75e-02 6.07e-02 0.62 0.537
## Year2002 -3.39e-03 5.88e-02 -0.06 0.954
## Year2003 -7.01e-02 6.18e-02 -1.14 0.256
## Year2004 1.02e-01 5.70e-02 1.79 0.073 .
## Year2005 1.00e-01 5.75e-02 1.75 0.081 .
```

```

## Year2006      -1.15e-02   5.85e-02   -0.20   0.844
## Year2007      3.21e-02   5.81e-02    0.55   0.581
## Year2008      5.46e-02   5.52e-02    0.99   0.323
## Year2009      1.13e-01   5.32e-02    2.13   0.033 *
## Year2010      3.17e-02   5.22e-02    0.61   0.544
## Year2011      4.29e-02   5.27e-02    0.82   0.415
## Year2012      1.31e-01   5.20e-02    2.52   0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.0091
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 281 is an outlier with |weight| = 0 ( < 4e-05);
## 204 weights are ~ = 1. The remaining 2287 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0384 0.8650 0.9500 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1          1.006
## Year              1.012 16          1.000

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2264 -0.2829 0.0173 0.2859 2.2240
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08068 0.04344 24.88 <2e-16 ***
## FirstAuthorFemale1 -0.00668 0.02220 -0.30 0.763
## Year1997 0.01846 0.05999 0.31 0.758
## Year1998 0.02729 0.06466 0.42 0.673
## Year1999 -0.01092 0.05937 -0.18 0.854
## Year2000 0.14572 0.05662 2.57 0.010 *
## Year2001 0.03755 0.06073 0.62 0.536
## Year2002 -0.00337 0.05884 -0.06 0.954
## Year2003 -0.07010 0.06175 -1.14 0.256
## Year2004 0.10205 0.05696 1.79 0.073 .
## Year2005 0.10049 0.05753 1.75 0.081 .
## Year2006 -0.01154 0.05849 -0.20 0.844
```

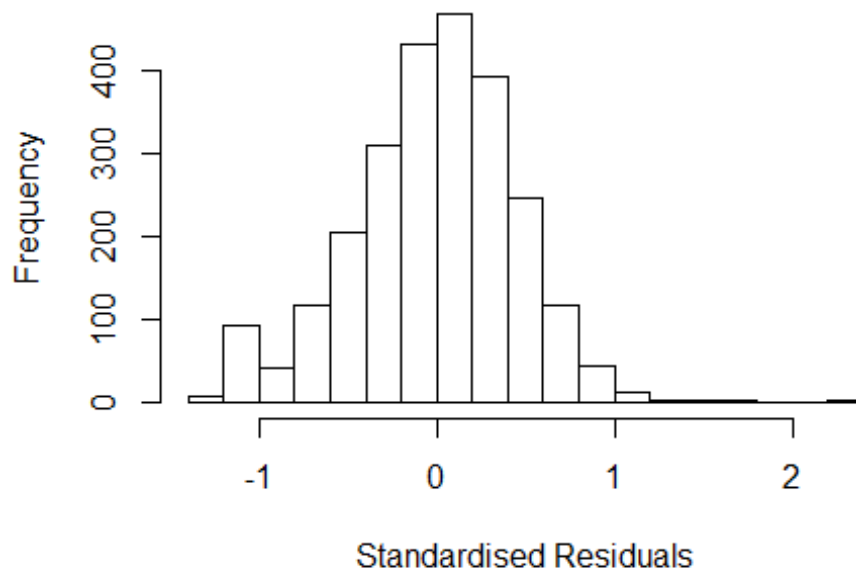


```

## Year2007          0.03210      0.05808      0.55      0.581
## Year2008          0.05460      0.05513      0.99      0.322
## Year2009          0.11336      0.05322      2.13      0.033 *
## Year2010          0.03170      0.05225      0.61      0.544
## Year2011          0.04293      0.05264      0.82      0.415
## Year2012          0.13124      0.05201      2.52      0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.0163, Adjusted R-squared:  0.0095
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 281 is an outlier with |weight| = 0 ( < 4e-05);
## 204 weights are ~= 1. The remaining 2287 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0381 0.8650 0.9500 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.01e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.03 1          1.015
## Year          1.03 16          1.001

```

## Residuals from last author



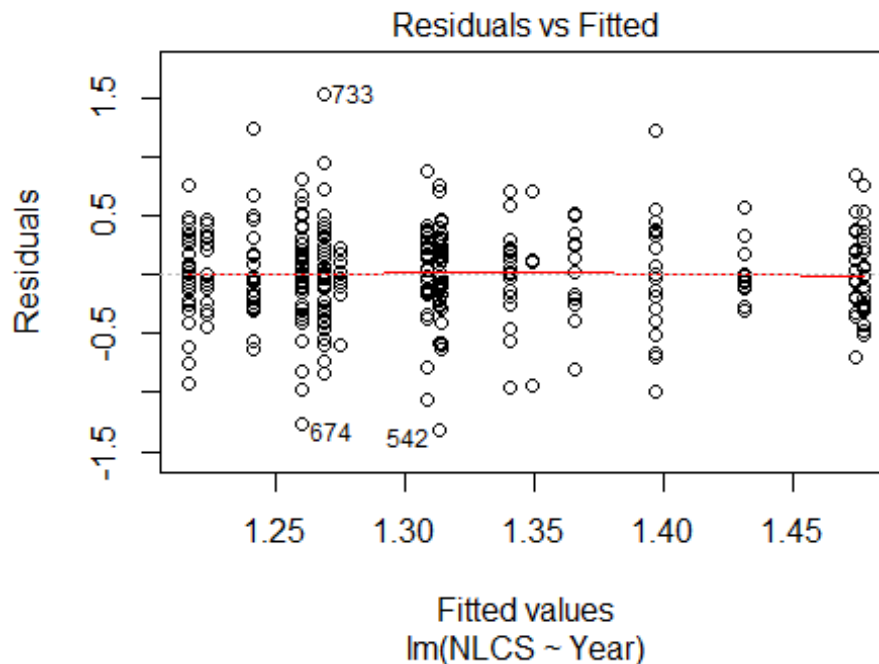
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2251 -0.2823 0.0156 0.2870 2.2252
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07943 0.04302 25.09 <2e-16 ***
## LastAuthorFemale1 -0.00152 0.02872 -0.05 0.958
## Year1997 0.01888 0.05996 0.31 0.753
## Year1998 0.02733 0.06463 0.42 0.672
## Year1999 -0.01046 0.05938 -0.18 0.860
## Year2000 0.14570 0.05661 2.57 0.010 *
## Year2001 0.03796 0.06066 0.63 0.531
## Year2002 -0.00297 0.05876 -0.05 0.960
## Year2003 -0.06984 0.06168 -1.13 0.258
## Year2004 0.10247 0.05688 1.80 0.072 .
## Year2005 0.10044 0.05754 1.75 0.081 .
## Year2006 -0.01127 0.05843 -0.19 0.847
```

```

## Year2007      0.03223    0.05804    0.56    0.579
## Year2008      0.05465    0.05515    0.99    0.322
## Year2009      0.11345    0.05321    2.13    0.033 *
## Year2010      0.03191    0.05217    0.61    0.541
## Year2011      0.04320    0.05261    0.82    0.412
## Year2012      0.13143    0.05196    2.53    0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.424
## Multiple R-squared:  0.0162, Adjusted R-squared:  0.00947
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 281 is an outlier with |weight| = 0 ( < 4e-05);
## 206 weights are ~= 1. The remaining 2285 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0401 0.8640 0.9500 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2492"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 1913"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   31   30   29   39   37   28   51   16   31   31   41   36   34   54   50
## 2011 2012
##   39   34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   15   13   18   19   4    31   11   21   19   26   25   21   39   26

```

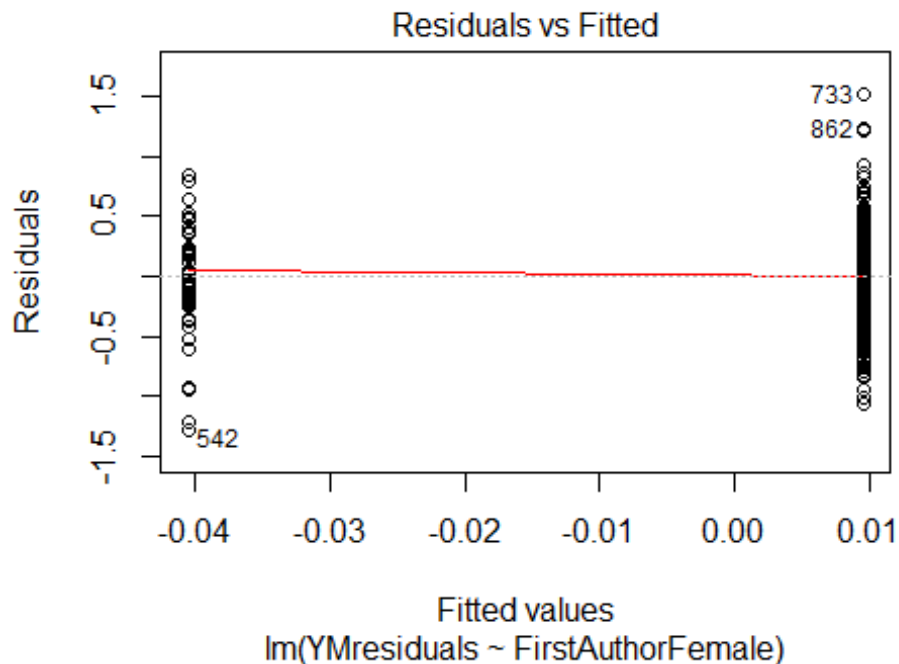
```
## 2011 2012
## 32 25
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 13 11 18 18 4 27 11 20 16 24 22 19 34 25
## 2011 2012
## 32 23
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.07
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.27, df = 1, p-value = 0.6
## [1] "Female first author team size 2018 geometric mean: 4.14060896224128"
## [1] "Male first author team size 2018 geometric mean: 2.86319182258129"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

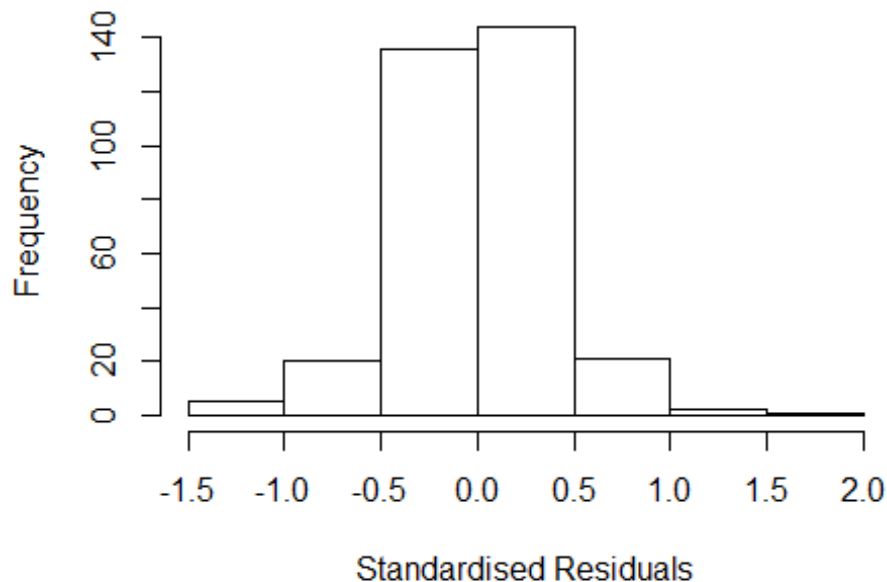
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 160, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.86636408986353"
## [1] "Male last author team size 2018 geometric mean: 3.00023477574853"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 96, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.407  1      1.186
## LastAuthorFemale  1.422  1      1.192
## UniqueAuthors    2.355  4      1.113
## Year              3.117 16      1.036
```

## Residuals from first and last author and team size



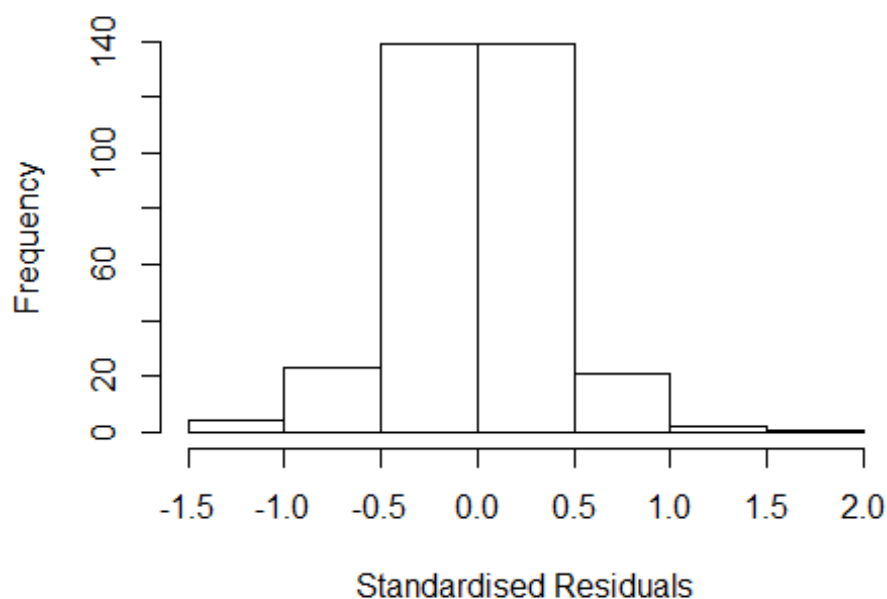
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20108 -0.22258 0.00979 0.21710 1.65865
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.34602 0.10014 13.44 <2e-16 ***
## FirstAuthorFemale1 -0.00206 0.05225 -0.04 0.969
## LastAuthorFemale1 -0.08990 0.06052 -1.49 0.138
## UniqueAuthors2 0.12654 0.06150 2.06 0.040 *
## UniqueAuthors3 0.16974 0.06797 2.50 0.013 *
## UniqueAuthors4 0.03776 0.08175 0.46 0.644
## UniqueAuthors5 0.17645 0.08822 2.00 0.046 *
## Year1997 -0.16207 0.17190 -0.94 0.347
## Year1998 0.02397 0.11968 0.20 0.841
## Year1999 0.06165 0.12346 0.50 0.618
```

```

## Year2000      -0.08707      0.15908      -0.55      0.585
## Year2001      0.16335      0.27941      0.58      0.559
## Year2002     -0.10085      0.11822     -0.85      0.394
## Year2003     -0.15569      0.11992     -1.30      0.195
## Year2004     -0.11098      0.12529     -0.89      0.376
## Year2005     -0.05448      0.12332     -0.44      0.659
## Year2006      0.04633      0.11743      0.39      0.693
## Year2007     -0.05298      0.11520     -0.46      0.646
## Year2008     -0.17635      0.11337     -1.56      0.121
## Year2009     -0.17902      0.11713     -1.53      0.127
## Year2010     -0.21167      0.11987     -1.77      0.078 .
## Year2011     -0.22047      0.10921     -2.02      0.044 *
## Year2012     -0.26423      0.12585     -2.10      0.037 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.339
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0399
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 254 is an outlier with |weight| = 0 ( < 0.0003);
## 26 weights are ~= 1. The remaining 302 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0674 0.8750 0.9540 0.8960 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.04e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.328 1 1.153
## LastAuthorFemale 1.321 1 1.149
## Year 1.470 16 1.012

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30251 -0.24600 -0.00625 0.22836 1.54118
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.44000 0.09669 14.89 <2e-16 ***
## FirstAuthorFemale1 0.00970 0.05258 0.18 0.854
## LastAuthorFemale1 -0.10249 0.05937 -1.73 0.085 .
## Year1997 -0.17737 0.18249 -0.97 0.332
## Year1998 0.00225 0.12065 0.02 0.985
## Year1999 0.02275 0.13195 0.17 0.863
## Year2000 -0.04906 0.17209 -0.29 0.776
## Year2001 0.11413 0.33372 0.34 0.733
## Year2002 -0.10671 0.12017 -0.89 0.375
## Year2003 -0.14767 0.12212 -1.21 0.227
## Year2004 -0.11737 0.12756 -0.92 0.358
## Year2005 -0.04433 0.12790 -0.35 0.729
```

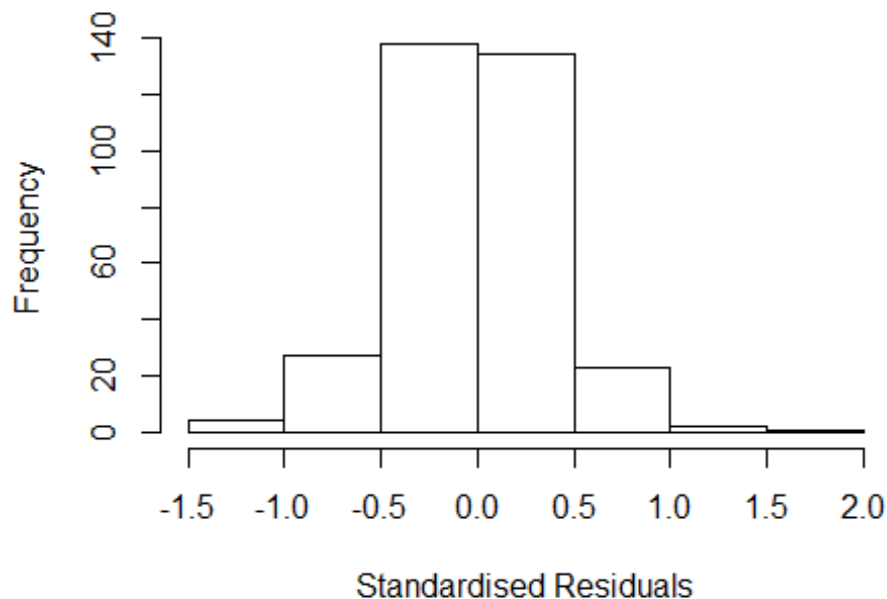


```

## Year2006          0.05575      0.11953      0.47      0.641
## Year2007          -0.04470      0.11896     -0.38      0.707
## Year2008          -0.16508      0.11610     -1.42      0.156
## Year2009          -0.17133      0.11889     -1.44      0.151
## Year2010          -0.18819      0.12140     -1.55      0.122
## Year2011          -0.20037      0.11424     -1.75      0.080 .
## Year2012          -0.22389      0.12315     -1.82      0.070 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.0692, Adjusted R-squared:  0.0151
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.026  0.884  0.949  0.899  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.246 1      1.116
## Year              1.246 16      1.007

```

## Residuals from first author



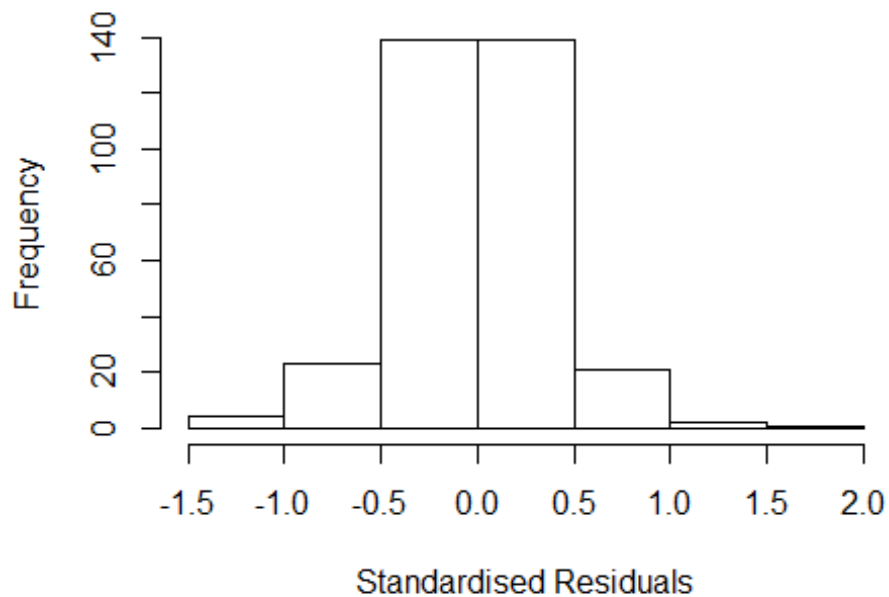
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3657 -0.2417 -0.0149 0.2319 1.5454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.43190 0.09259 15.46 <2e-16 ***
## FirstAuthorFemale1 -0.01739 0.05127 -0.34 0.735
## Year1997 -0.17874 0.18101 -0.99 0.324
## Year1998 0.00289 0.11782 0.02 0.980
## Year1999 0.03146 0.12970 0.24 0.809
## Year2000 -0.05232 0.17038 -0.31 0.759
## Year2001 0.07437 0.34574 0.22 0.830
## Year2002 -0.11630 0.11625 -1.00 0.318
## Year2003 -0.13877 0.11682 -1.19 0.236
## Year2004 -0.11991 0.12278 -0.98 0.330
## Year2005 -0.03311 0.12427 -0.27 0.790
## Year2006 0.05709 0.11652 0.49 0.625
```

```

## Year2007          -0.04879    0.11529   -0.42    0.672
## Year2008          -0.17739    0.11453   -1.55    0.122
## Year2009          -0.17277    0.11574   -1.49    0.137
## Year2010          -0.18432    0.11853   -1.56    0.121
## Year2011          -0.19367    0.11088   -1.75    0.082 .
## Year2012          -0.23243    0.12129   -1.92    0.056 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.363
## Multiple R-squared:  0.0615, Adjusted R-squared:  0.0102
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0303 0.8850 0.9520 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.232 1      1.110
## Year              1.232 16      1.007

```

## Residuals from last author



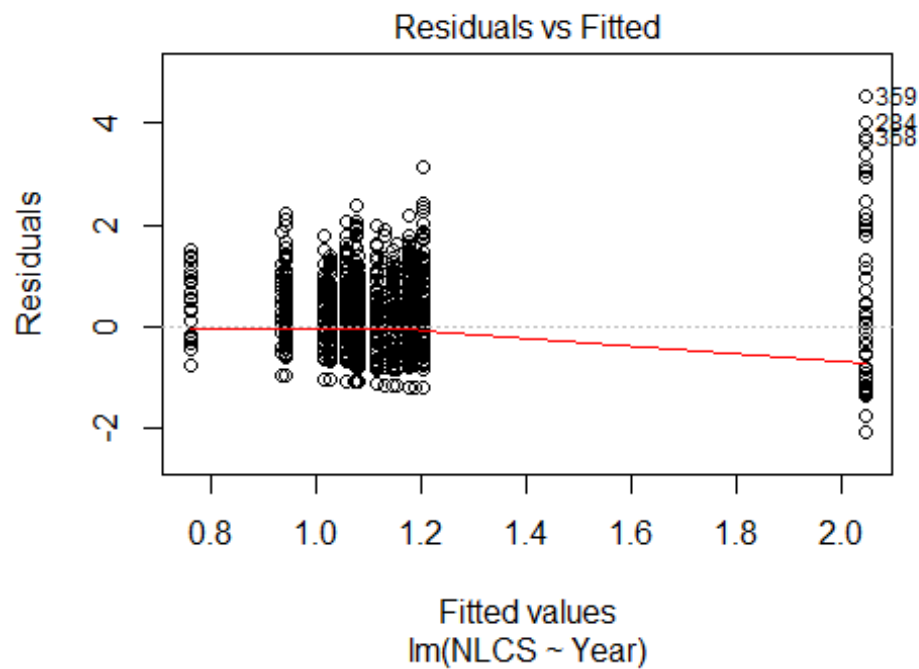
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29712 -0.24783 -0.00458 0.22862 1.53926
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.43982 0.09640 14.94 <2e-16 ***
## LastAuthorFemale1 -0.09954 0.05758 -1.73 0.085 .
## Year1997 -0.17594 0.18196 -0.97 0.334
## Year1998 0.00328 0.12026 0.03 0.978
## Year1999 0.02513 0.13106 0.19 0.848
## Year2000 -0.04849 0.17176 -0.28 0.778
## Year2001 0.11310 0.33212 0.34 0.734
## Year2002 -0.10474 0.11916 -0.88 0.380
## Year2003 -0.14424 0.11966 -1.21 0.229
## Year2004 -0.11744 0.12709 -0.92 0.356
## Year2005 -0.04305 0.12750 -0.34 0.736
## Year2006 0.05799 0.11877 0.49 0.626
```

```

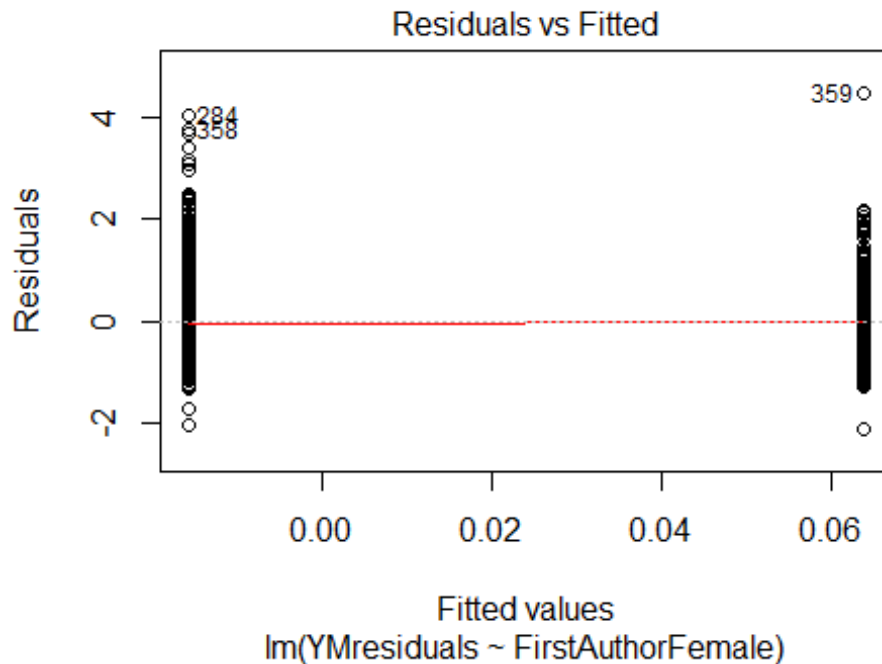
## Year2007          -0.04315      0.11812    -0.37      0.715
## Year2008          -0.16264      0.11560    -1.41      0.160
## Year2009          -0.17136      0.11877    -1.44      0.150
## Year2010          -0.18608      0.12049    -1.54      0.124
## Year2011          -0.19824      0.11355    -1.75      0.082 .
## Year2012          -0.22099      0.12268    -1.80      0.073 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.362
## Multiple R-squared:  0.0688, Adjusted R-squared:  0.0179
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0319 0.8870 0.9500 0.9010 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 329"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2000"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   59   55   71   70  107   91  109   97  111  136  142  205  194  243  227
## 2011 2012
##  236  216
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   45   39   53   57   79   72   88   78   82  113  120  171  154  204  180
## 2011 2012

```

```
## 190 170
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 39 50 57 75 70 82 70 77 106 114 166 146 188 173
## 2011 2012
## 179 163
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 260, df = 16, p-value <2e-16
```

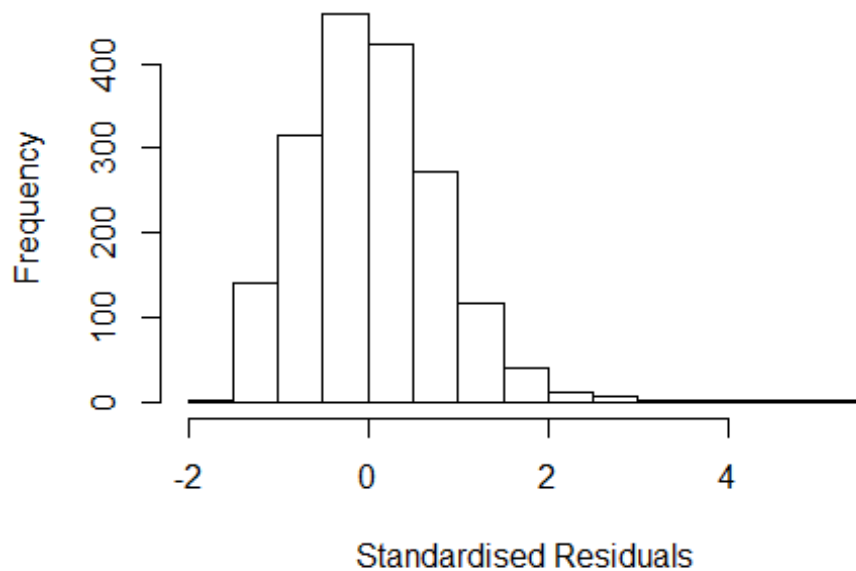


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00022, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.4572738192877"
## [1] "Male first author team size 2018 geometric mean: 1.43638196557519"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.51911245999068"
## [1] "Male last author team size 2018 geometric mean: 1.41150214736178"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.759 1          1.326
## LastAuthorFemale  1.748 1          1.322
## UniqueAuthors    1.182 4          1.021
## Year             1.244 16          1.007
```

## Residuals from first and last author and team size



```
## [1] "List of 19 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 280   0001712947 4.165 2000     2000      1    3.036
## 284   4444330855 6.077 2000     2000      1    4.948
## 330   0001811888 5.125 2000     2000      1    3.790
## 331   0004467257 3.945 2000     2000      1    2.713
## 333   0012604466 4.264 2000     2000      1    2.710
## 334   0141884822 5.179 2000     2000      1    3.844
## 336   33845353104 3.945 2000     2000      1    2.816
## 340   84990342713 4.528 2000     2000      1    3.177
## 341   84990353485 4.059 2000     2000      1    2.607
## 342   84990354003 3.822 2000     2000      1    2.693
## 344   84990374547 3.945 2000     2000      1    2.713
## 345   84992784725 5.711 2000     2000      1    4.376
## 346   84992793050 5.423 2000     2000      1    4.088
## 352   84992804274 5.011 2000     2000      1    3.882
## 358   84992875320 5.785 2000     2000      1    4.334
## 359   84992885731 6.577 2000     2000      1    5.345
## 361   84997860281 4.165 2000     2000      1    2.742
## 2793  84856185034 4.364 2012     2000      1    3.026
## 2818  84863845958 3.555 2012     2000      1    2.539
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
```



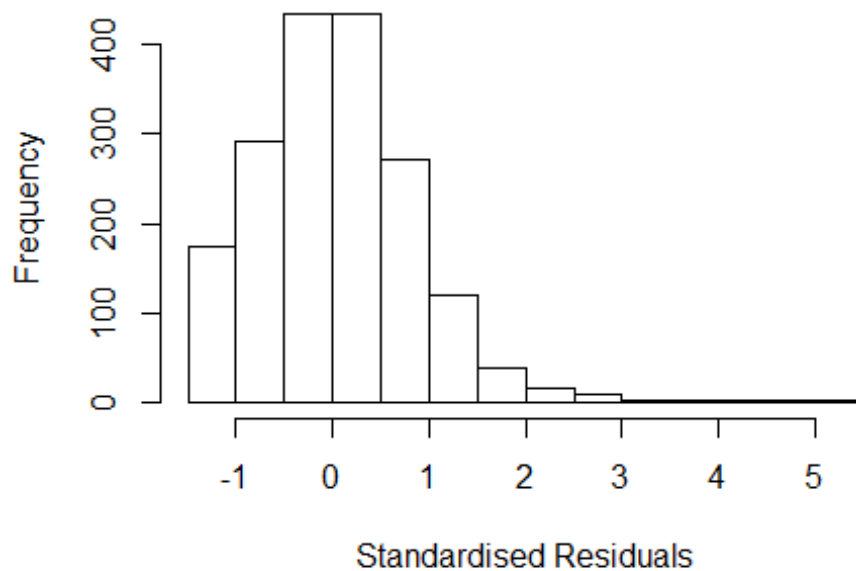
```

##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q        Max
## -1.53857 -0.50843 -0.00663  0.51970  5.34511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6561    0.1147    5.72 1.2e-08 ***
## FirstAuthorFemale1 0.0874    0.0599    1.46  0.1443
## LastAuthorFemale1 0.0155    0.0577    0.27  0.7887
## UniqueAuthors2    0.2065    0.0440    4.69 2.9e-06 ***
## UniqueAuthors3    0.3221    0.0643    5.01 6.0e-07 ***
## UniqueAuthors4    0.3227    0.1124    2.87  0.0041 **
## UniqueAuthors5    0.1584    0.2207    0.72  0.4728
## Year1997          0.1599    0.1935    0.83  0.4085
## Year1998          0.3043    0.1539    1.98  0.0482 *
## Year1999          0.4054    0.1508    2.69  0.0072 **
## Year2000          0.4729    0.2818    1.68  0.0935 .
## Year2001          0.3901    0.1487    2.62  0.0088 **
## Year2002          0.3473    0.1402    2.48  0.0133 *
## Year2003          0.1859    0.1407    1.32  0.1867
## Year2004          0.2905    0.1480    1.96  0.0498 *
## Year2005          0.2121    0.1299    1.63  0.1025
## Year2006          0.3155    0.1290    2.45  0.0145 *
## Year2007          0.3155    0.1260    2.50  0.0124 *
## Year2008          0.2439    0.1323    1.84  0.0655 .
## Year2009          0.1497    0.1244    1.20  0.2291
## Year2010          0.2808    0.1278    2.20  0.0282 *
## Year2011          0.3992    0.1256    3.18  0.0015 **
## Year2012          0.3598    0.1331    2.70  0.0069 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.753
## Multiple R-squared:  0.0453, Adjusted R-squared:  0.0335
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
##  8 observations c(195,219,223,232,233,237,241,242)
##  are outliers with |weight| = 0 ( < 5.6e-05);
##  160 weights are ~ = 1. The remaining 1632 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0353 0.8670 0.9500 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw

```

```
##          5.00e-01          5.00e-01
##  nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##        500         50         2         1        1000         200
##  trace.lev      mts    compute.rd
##        0         1000         0
##          psi      subsampling      cov
##    "bisquare"    "nonsingular"    ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.757 1          1.326
## LastAuthorFemale  1.722 1          1.312
## Year              1.080 16          1.002
```

### Residuals from first and last author



```
## [1] "List of 18 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 280  0001712947 4.165 2000    2000      1    2.904
## 284  4444330855 6.077 2000    2000      1    4.816
## 330  0001811888 5.125 2000    2000      1    3.864
## 331  0004467257 3.945 2000    2000      1    2.563
## 333  0012604466 4.264 2000    2000      1    2.882
## 334  0141884822 5.179 2000    2000      1    3.918
## 336  33845353104 3.945 2000    2000      1    2.684
## 340  84990342713 4.528 2000    2000      1    3.228
```

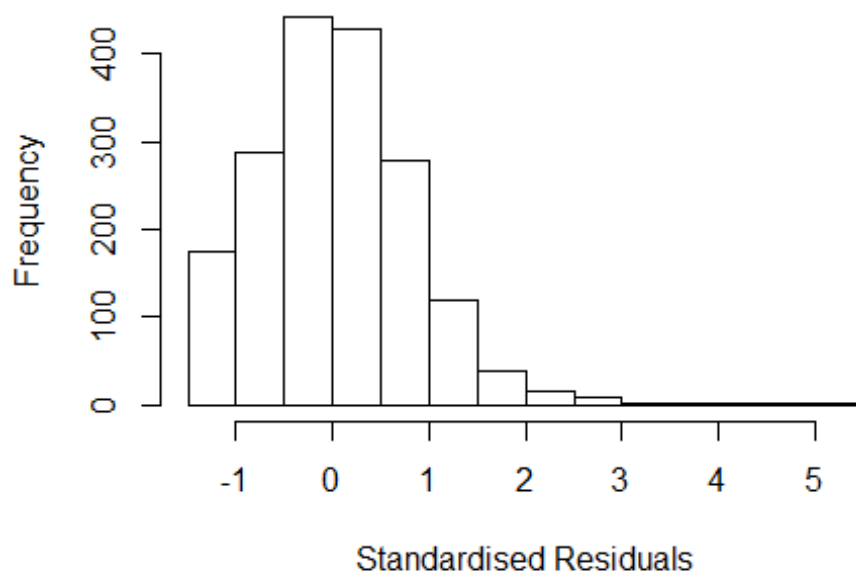
```

## 341 84990353485 4.059 2000      2000      1      2.798
## 342 84990354003 3.822 2000      2000      1      2.561
## 344 84990374547 3.945 2000      2000      1      2.563
## 345 84992784725 5.711 2000      2000      1      4.450
## 346 84992793050 5.423 2000      2000      1      4.162
## 352 84992804274 5.011 2000      2000      1      3.750
## 358 84992875320 5.785 2000      2000      1      4.524
## 359 84992885731 6.577 2000      2000      1      5.195
## 361 84997860281 4.165 2000      2000      1      2.822
## 2793 84856185034 4.364 2012      2000      1      3.283
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38197 -0.51413 -0.00285  0.51112  5.19503
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7150     0.1153   6.20 6.9e-10 ***
## FirstAuthorFemale1 0.0816     0.0597   1.37  0.17208
## LastAuthorFemale1 0.0394     0.0574   0.69  0.49231
## Year1997          0.1812     0.1973   0.92  0.35843
## Year1998          0.3216     0.1534   2.10  0.03624 *
## Year1999          0.4151     0.1509   2.75  0.00600 **
## Year2000          0.5459     0.2854   1.91  0.05593 .
## Year2001          0.4410     0.1513   2.91  0.00361 **
## Year2002          0.3867     0.1399   2.76  0.00577 **
## Year2003          0.2120     0.1439   1.47  0.14074
## Year2004          0.3192     0.1494   2.14  0.03282 *
## Year2005          0.2439     0.1317   1.85  0.06420 .
## Year2006          0.3474     0.1312   2.65  0.00819 **
## Year2007          0.3352     0.1275   2.63  0.00864 **
## Year2008          0.2569     0.1333   1.93  0.05411 .
## Year2009          0.1609     0.1259   1.28  0.20153
## Year2010          0.2944     0.1295   2.27  0.02318 *
## Year2011          0.4212     0.1276   3.30  0.00098 ***
## Year2012          0.3662     0.1342   2.73  0.00643 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.768
## Multiple R-squared:  0.022, Adjusted R-squared:  0.0121
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 8 observations c(195,219,223,232,233,237,241,242)

```

```
## are outliers with |weight| = 0 ( < 5.6e-05);
## 127 weights are ~= 1. The remaining 1665 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0279 0.8690 0.9530 0.9110 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.06 1      1.030
## Year      1.06 16      1.002
```

## Residuals from first author



```
## [1] "List of 18 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
```

```

## 280 0001712947 4.165 2000 2000 1 2.904
## 284 4444330855 6.077 2000 2000 1 4.816
## 330 0001811888 5.125 2000 2000 1 3.864
## 331 0004467257 3.945 2000 2000 1 2.563
## 333 0012604466 4.264 2000 2000 1 2.882
## 334 0141884822 5.179 2000 2000 1 3.918
## 336 33845353104 3.945 2000 2000 1 2.684
## 340 84990342713 4.528 2000 2000 1 3.228
## 341 84990353485 4.059 2000 2000 1 2.798
## 342 84990354003 3.822 2000 2000 1 2.561
## 344 84990374547 3.945 2000 2000 1 2.563
## 345 84992784725 5.711 2000 2000 1 4.450
## 346 84992793050 5.423 2000 2000 1 4.162
## 352 84992804274 5.011 2000 2000 1 3.750
## 358 84992875320 5.785 2000 2000 1 4.524
## 359 84992885731 6.577 2000 2000 1 5.195
## 361 84997860281 4.165 2000 2000 1 2.822
## 2793 84856185034 4.364 2012 2000 1 3.283
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37066 -0.51172 -0.00555 0.51935 5.20634
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7193 0.1156 6.22 6.1e-10 ***
## FirstAuthorFemale1 0.1074 0.0464 2.32 0.0206 *
## Year1997 0.1778 0.1973 0.90 0.3676
## Year1998 0.3217 0.1533 2.10 0.0360 *
## Year1999 0.4124 0.1512 2.73 0.0065 **
## Year2000 0.5439 0.2854 1.91 0.0569 .
## Year2001 0.4403 0.1517 2.90 0.0037 **
## Year2002 0.3834 0.1403 2.73 0.0063 **
## Year2003 0.2113 0.1442 1.47 0.1431
## Year2004 0.3188 0.1500 2.13 0.0337 *
## Year2005 0.2419 0.1319 1.83 0.0669 .
## Year2006 0.3468 0.1315 2.64 0.0084 **
## Year2007 0.3339 0.1278 2.61 0.0090 **
## Year2008 0.2564 0.1337 1.92 0.0553 .
## Year2009 0.1589 0.1263 1.26 0.2084
## Year2010 0.2942 0.1298 2.27 0.0236 *
## Year2011 0.4209 0.1278 3.29 0.0010 **
## Year2012 0.3663 0.1344 2.73 0.0065 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

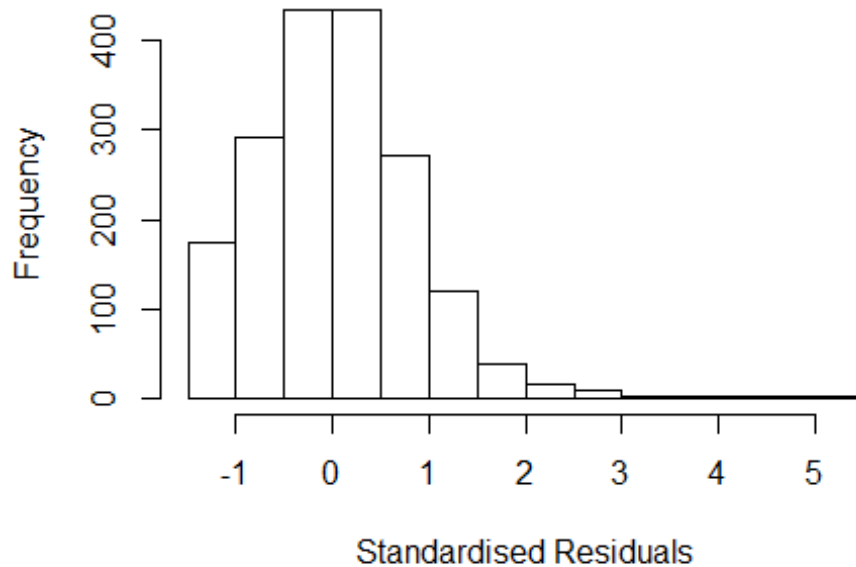
```

```

##
## Robust residual standard error: 0.768
## Multiple R-squared: 0.0217, Adjusted R-squared: 0.0124
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 8 observations c(195,219,223,232,233,237,241,242)
## are outliers with |weight| = 0 ( < 5.6e-05);
## 135 weights are ~= 1. The remaining 1657 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0291 0.8670 0.9520 0.9110 0.9860 0.9990
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 5.56e-05 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 1000 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.039 1 1.019
## Year 1.039 16 1.001

```

## Residuals from last author



```
## [1] "List of 18 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 280   0001712947 4.165 2000     2000      1    2.904
## 284   4444330855 6.077 2000     2000      1    4.816
## 330   0001811888 5.125 2000     2000      1    3.864
## 331   0004467257 3.945 2000     2000      1    2.563
## 333   0012604466 4.264 2000     2000      1    2.882
## 334   0141884822 5.179 2000     2000      1    3.918
## 336   33845353104 3.945 2000     2000      1    2.684
## 340   84990342713 4.528 2000     2000      1    3.228
## 341   84990353485 4.059 2000     2000      1    2.798
## 342   84990354003 3.822 2000     2000      1    2.561
## 344   84990374547 3.945 2000     2000      1    2.563
## 345   84992784725 5.711 2000     2000      1    4.450
## 346   84992793050 5.423 2000     2000      1    4.162
## 352   84992804274 5.011 2000     2000      1    3.750
## 358   84992875320 5.785 2000     2000      1    4.524
## 359   84992885731 6.577 2000     2000      1    5.195
## 361   84997860281 4.165 2000     2000      1    2.822
## 2793  84856185034 4.364 2012     2000      1    3.283
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
```

```

##      Min      1Q   Median      3Q      Max
## -1.35910 -0.52645 -0.00289  0.52122  5.21790
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7130      0.1143   6.24 5.6e-10 ***
## LastAuthorFemale1 0.0901      0.0445   2.03  0.0430 *
## Year1997          0.1869      0.1974   0.95  0.3440
## Year1998          0.3254      0.1531   2.13  0.0336 *
## Year1999          0.4222      0.1492   2.83  0.0047 **
## Year2000          0.5560      0.2831   1.96  0.0497 *
## Year2001          0.4455      0.1503   2.96  0.0031 **
## Year2002          0.3962      0.1384   2.86  0.0043 **
## Year2003          0.2218      0.1427   1.55  0.1203
## Year2004          0.3256      0.1482   2.20  0.0281 *
## Year2005          0.2537      0.1307   1.94  0.0523 .
## Year2006          0.3535      0.1302   2.72  0.0067 **
## Year2007          0.3455      0.1264   2.73  0.0063 **
## Year2008          0.2645      0.1323   2.00  0.0456 *
## Year2009          0.1734      0.1246   1.39  0.1643
## Year2010          0.2999      0.1285   2.33  0.0198 *
## Year2011          0.4252      0.1266   3.36  0.0008 ***
## Year2012          0.3755      0.1335   2.81  0.0050 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.769
## Multiple R-squared:  0.0209, Adjusted R-squared:  0.0115
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 8 observations c(195,219,223,232,233,237,241,242)
## are outliers with |weight| = 0 ( < 5.6e-05);
## 136 weights are ~= 1. The remaining 1656 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0304 0.8660 0.9520 0.9110 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.56e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats

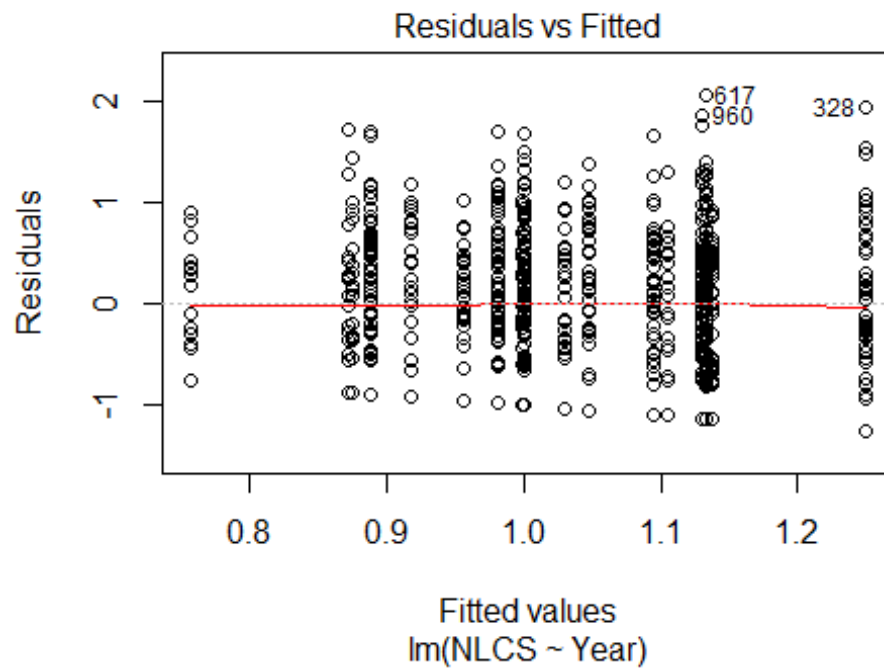
```



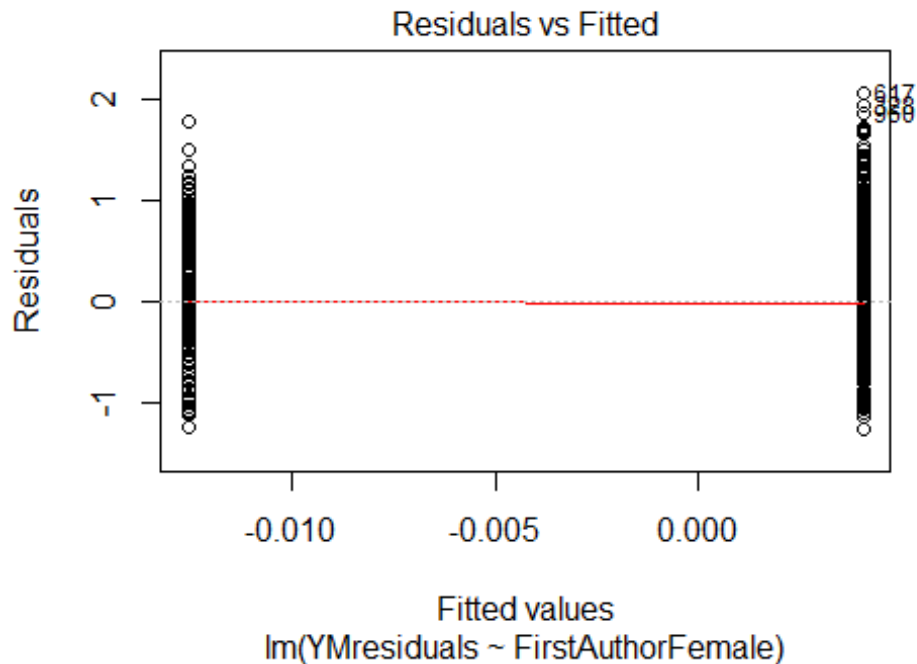
```

##                                "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1800"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2001"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    25    26    26    24    35    39    27    46    39    59    53    74   110   120   116
## 2011 2012
##   102   134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    21    22    23    19    31    33    20    41    35    48    43    59    99   104   100
## 2011 2012
##    91   110
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    20    22    23    19    30    32    19    39    34    45    40    55    92    95    97
## 2011 2012
##    86   103
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 10, df = 16, p-value = 0.9

```

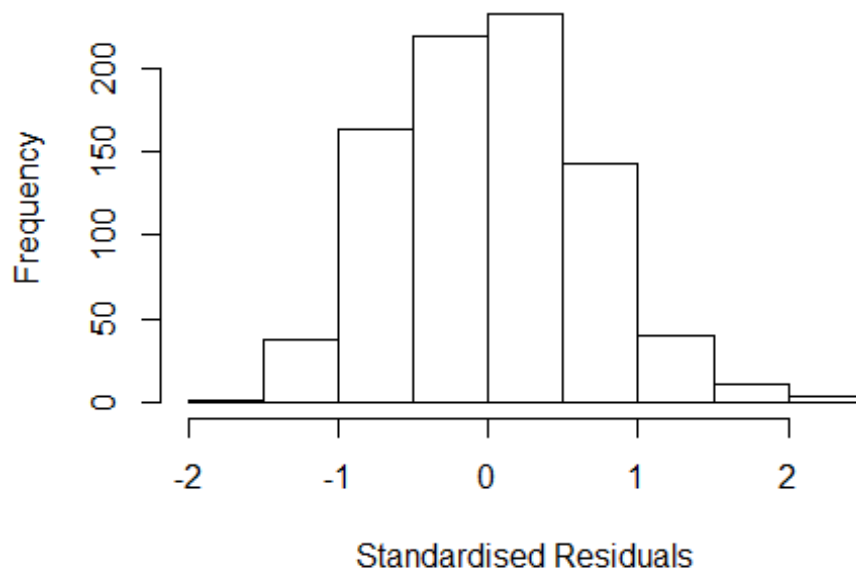


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.083, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 1.47337053278629"
## [1] "Male first author team size 2018 geometric mean: 1.30963609299372"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.32173502924764"
## [1] "Male last author team size 2018 geometric mean: 1.39059515127844"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.582 1          1.258
## LastAuthorFemale  1.638 1          1.280
## UniqueAuthors    1.468 4          1.049
## Year              1.371 16         1.010
```

## Residuals from first and last author and team size



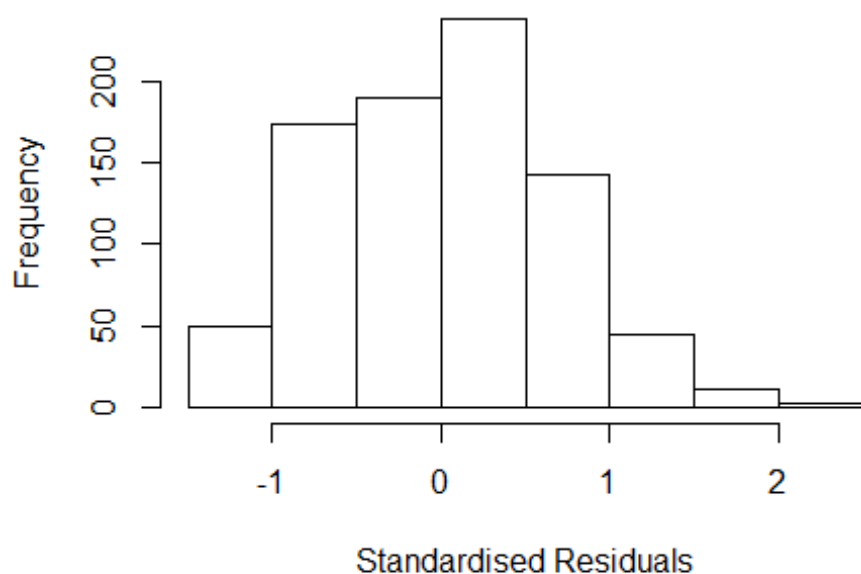
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5896 -0.4387 0.0184 0.4568 2.1454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8510 0.1663 5.12 3.8e-07 ***
## FirstAuthorFemale1 -0.0769 0.0667 -1.15 0.2493
## LastAuthorFemale1 0.0466 0.0678 0.69 0.4922
## UniqueAuthors2 0.2699 0.0586 4.61 4.7e-06 ***
## UniqueAuthors3 0.3207 0.0634 5.06 5.1e-07 ***
## UniqueAuthors4 0.5180 0.1701 3.05 0.0024 **
## UniqueAuthors5 0.3739 0.1380 2.71 0.0069 **
## Year1997 -0.0596 0.2119 -0.28 0.7787
## Year1998 -0.1105 0.2079 -0.53 0.5952
## Year1999 -0.2368 0.2013 -1.18 0.2398
```

```

## Year2000          0.1180      0.2041      0.58      0.5635
## Year2001         -0.0737      0.1978     -0.37      0.7097
## Year2002          0.1403      0.2170      0.65      0.5179
## Year2003          0.2228      0.1948      1.14      0.2532
## Year2004          0.0498      0.1873      0.27      0.7902
## Year2005          0.2206      0.2035      1.08      0.2787
## Year2006          0.0168      0.1966      0.09      0.9319
## Year2007          0.1685      0.1866      0.90      0.3667
## Year2008          0.1876      0.1835      1.02      0.3068
## Year2009         -0.0408      0.1759     -0.23      0.8165
## Year2010          0.1305      0.1751      0.75      0.4564
## Year2011         -0.0965      0.1779     -0.54      0.5877
## Year2012          0.0013      0.1758      0.01      0.9941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.691
## Multiple R-squared:  0.0737, Adjusted R-squared:  0.0491
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 782 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.315  0.879  0.952  0.918  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.18e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.608 1          1.268
## LastAuthorFemale  1.603 1          1.266
## Year              1.113 16          1.003

```

## Residuals from first and last author



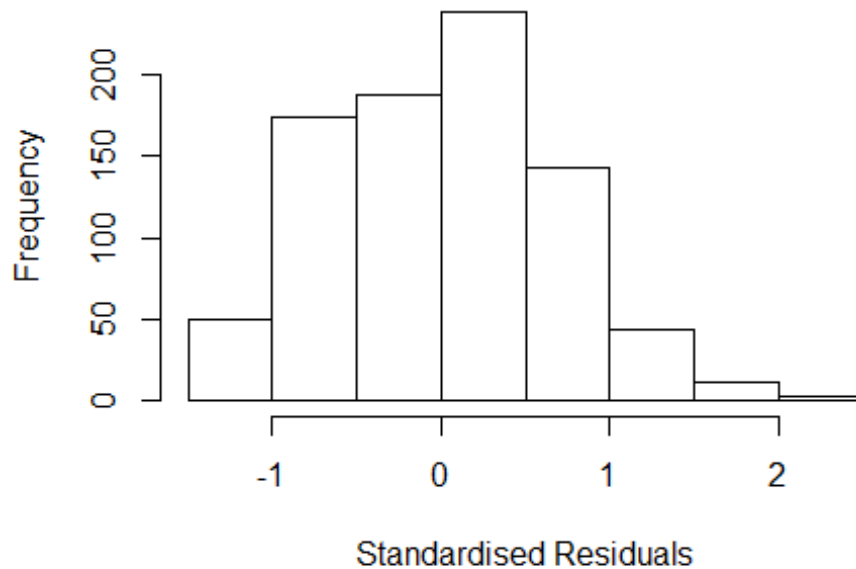
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1855 -0.5147 0.0316 0.4709 2.0499
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98595 0.16776 5.88 6e-09 ***
## FirstAuthorFemale1 -0.02957 0.06994 -0.42 0.67
## LastAuthorFemale1 0.03499 0.07061 0.50 0.62
## Year1997 -0.12718 0.22663 -0.56 0.57
## Year1998 -0.15809 0.21266 -0.74 0.46
## Year1999 -0.23224 0.21087 -1.10 0.27
## Year2000 0.06503 0.20457 0.32 0.75
## Year2001 -0.11028 0.21088 -0.52 0.60
## Year2002 0.07692 0.22230 0.35 0.73
## Year2003 0.16486 0.19933 0.83 0.41
## Year2004 -0.01413 0.19135 -0.07 0.94
## Year2005 0.19417 0.21018 0.92 0.36
```

```

## Year2006      0.03755    0.20982    0.18    0.86
## Year2007      0.14435    0.19206    0.75    0.45
## Year2008      0.14819    0.18574    0.80    0.43
## Year2009     -0.06800    0.18133   -0.38    0.71
## Year2010      0.12169    0.17959    0.68    0.50
## Year2011     -0.09923    0.18353   -0.54    0.59
## Year2012     -0.00458    0.18008   -0.03    0.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.0261, Adjusted R-squared:  0.00499
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 787 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.385  0.870  0.951  0.919  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.18e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.049 1      1.024
## Year              1.049 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1827 -0.5172 0.0294 0.4661 2.0482
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99180 0.16576 5.98 3.2e-09 ***
## FirstAuthorFemale1 -0.00781 0.05642 -0.14 0.89
## Year1997 -0.13209 0.22508 -0.59 0.56
## Year1998 -0.15615 0.21166 -0.74 0.46
## Year1999 -0.23148 0.21073 -1.10 0.27
## Year2000 0.06038 0.20346 0.30 0.77
## Year2001 -0.11356 0.20969 -0.54 0.59
## Year2002 0.07132 0.22009 0.32 0.75
## Year2003 0.15873 0.19717 0.81 0.42
## Year2004 -0.01590 0.19002 -0.08 0.93
## Year2005 0.19090 0.20892 0.91 0.36
## Year2006 0.03368 0.20831 0.16 0.87
```

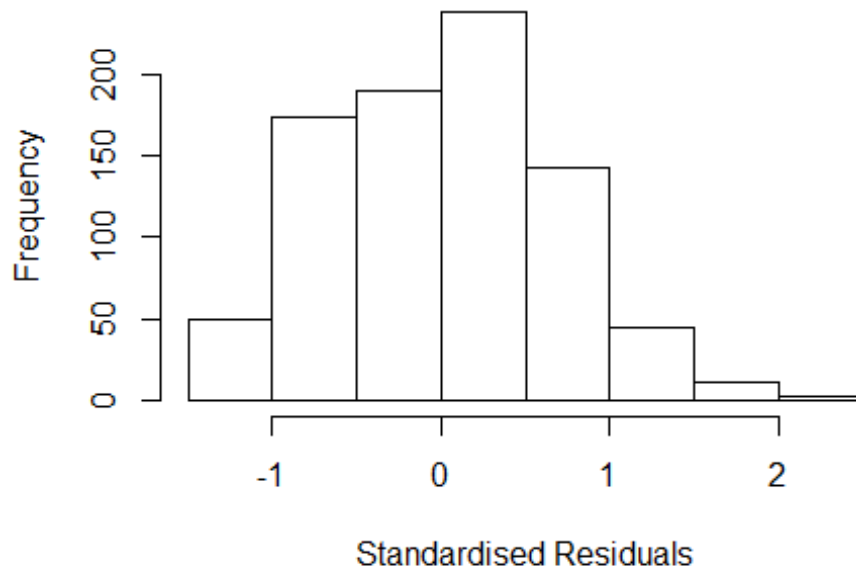


```

## Year2007          0.14112      0.19059      0.74      0.46
## Year2008          0.14397      0.18447      0.78      0.44
## Year2009         -0.06981      0.18004     -0.39      0.70
## Year2010          0.11885      0.17812      0.67      0.50
## Year2011         -0.10301      0.18166     -0.57      0.57
## Year2012         -0.00767      0.17856     -0.04      0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.714
## Multiple R-squared:  0.0258, Adjusted R-squared:  0.00588
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 784 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.391  0.870   0.951   0.919   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.18e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.05 1          1.025
## Year              1.05 16          1.002

```

## Residuals from last author



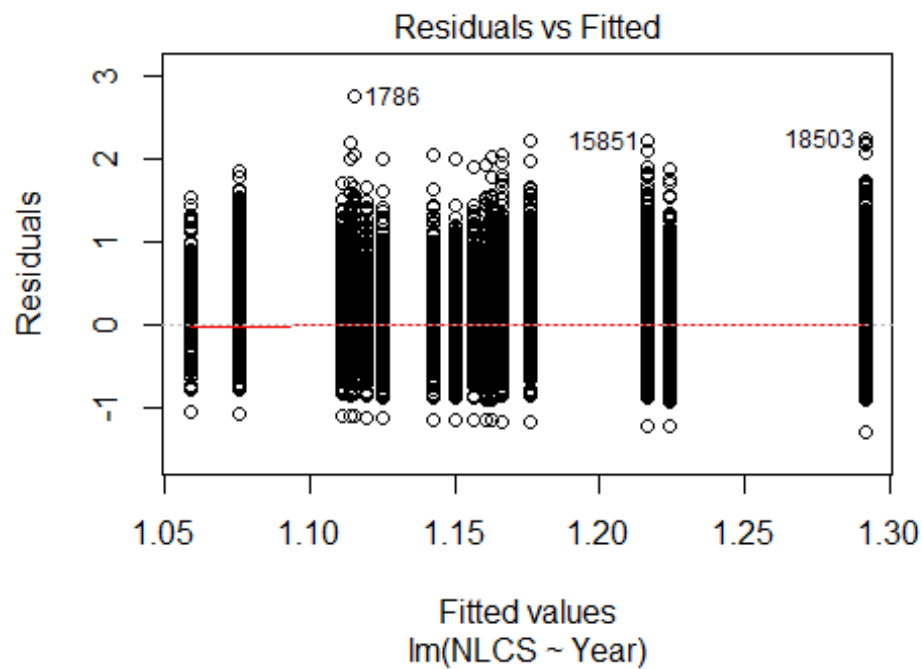
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1922 -0.5141 0.0347 0.4682 2.0523
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98431 0.16732 5.88 5.8e-09 ***
## LastAuthorFemale1 0.01551 0.05757 0.27 0.79
## Year1997 -0.12759 0.22597 -0.56 0.57
## Year1998 -0.15732 0.21268 -0.74 0.46
## Year1999 -0.23031 0.21107 -1.09 0.28
## Year2000 0.06198 0.20393 0.30 0.76
## Year2001 -0.10925 0.21028 -0.52 0.60
## Year2002 0.07309 0.22170 0.33 0.74
## Year2003 0.16211 0.19854 0.82 0.41
## Year2004 -0.01410 0.19103 -0.07 0.94
## Year2005 0.19240 0.20965 0.92 0.36
## Year2006 0.03593 0.20940 0.17 0.86
```

```

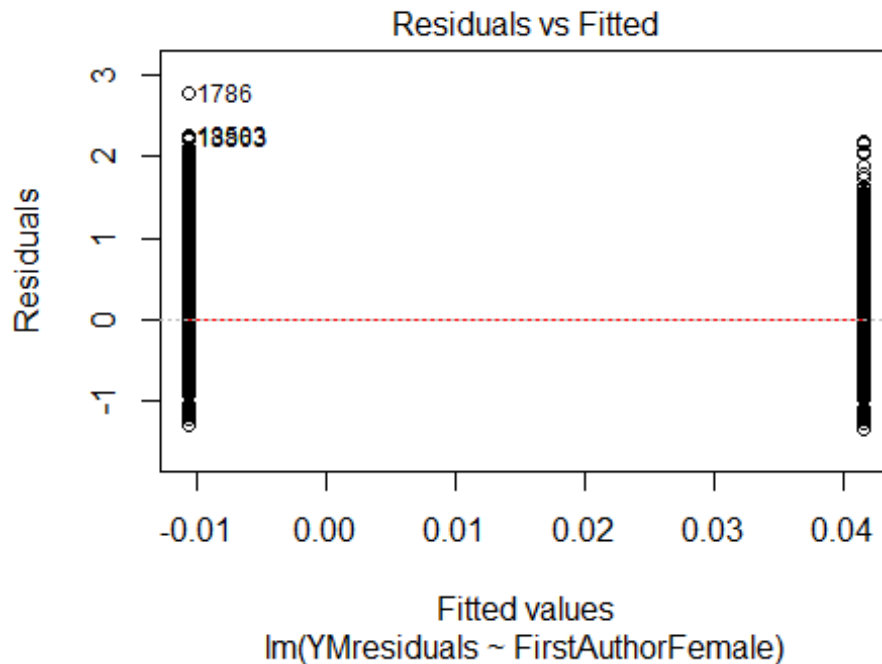
## Year2007      0.14488      0.19183      0.76      0.45
## Year2008      0.14738      0.18527      0.80      0.43
## Year2009     -0.06879      0.18086     -0.38      0.70
## Year2010      0.12008      0.17904      0.67      0.50
## Year2011     -0.10177      0.18282     -0.56      0.58
## Year2012     -0.00653      0.17946     -0.04      0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.709
## Multiple R-squared:  0.0259, Adjusted R-squared:  0.00603
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 789 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.382  0.869  0.951  0.919  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.18e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 851"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2002"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   712  691  708  731  839  560  891  669  725  763  810  931  943 1182 1140
## 2011 2012
## 1109 1178
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   567  540  549  564  607  348  721  539  590  629  660  745  758  935  934
## 2011 2012

```

```
## 912 969
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 551 522 523 544 565 330 688 508 560 604 608 683 704 878 872
## 2011 2012
## 842 893
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 160, df = 16, p-value <2e-16
```

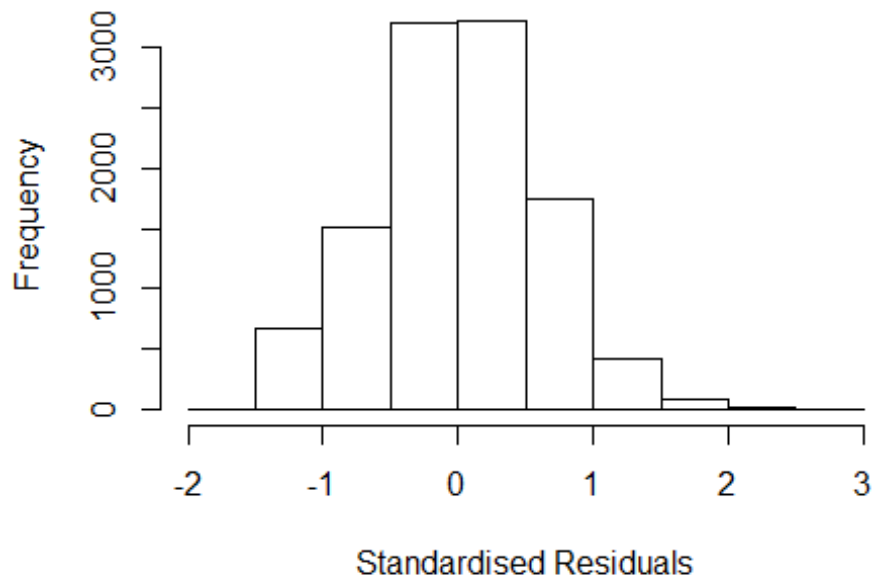


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.63021241923701"
## [1] "Male first author team size 2018 geometric mean: 1.53014961493491"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.53223327153767"
## [1] "Male last author team size 2018 geometric mean: 1.56890446330791"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 62000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.433  1          1.197
## LastAuthorFemale  1.422  1          1.193
## UniqueAuthors    1.043  4          1.005
## Year              1.046 16          1.001
```

## Residuals from first and last author and team size



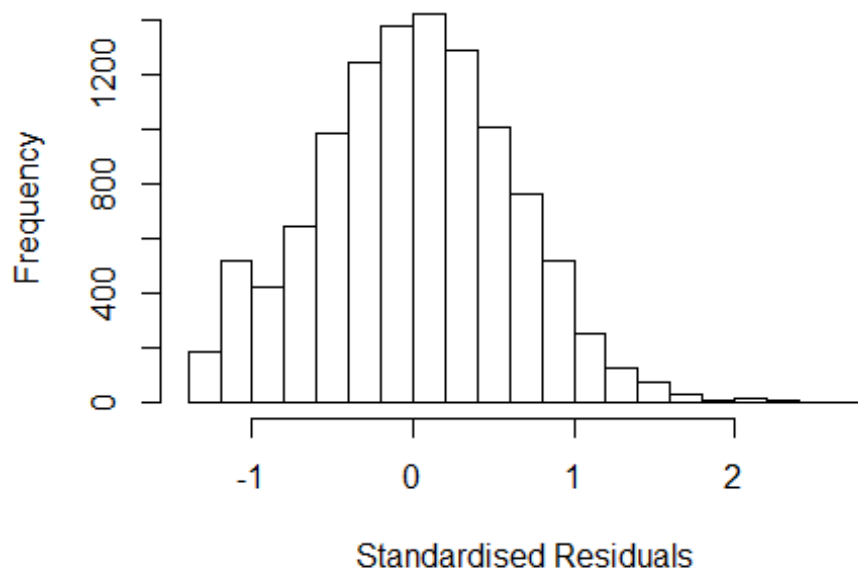
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1786 0001438979 3.878 1998      2002      2      2.718
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.54871 -0.40190  0.00475  0.41067  2.71827
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.99163    0.02856   34.72 < 2e-16 ***
## FirstAuthorFemale1 0.00219    0.01750    0.13  0.9002
## LastAuthorFemale1 0.07882    0.01762    4.47 7.8e-06 ***
## UniqueAuthors2    0.12848    0.01349    9.52 < 2e-16 ***
## UniqueAuthors3    0.21416    0.01935   11.07 < 2e-16 ***
## UniqueAuthors4    0.19575    0.03754    5.21 1.9e-07 ***
## UniqueAuthors5    0.43539    0.06346    6.86 7.2e-12 ***
## Year1997          0.05018    0.03821    1.31  0.1891
## Year1998          0.03962    0.03932    1.01  0.3137
## Year1999          0.10078    0.03845    2.62  0.0088 **
```

```

## Year2000          0.03709      0.03687      1.01      0.3145
## Year2001         -0.02345      0.04182     -0.56      0.5750
## Year2002          0.02100      0.03596      0.58      0.5593
## Year2003          0.07440      0.03833      1.94      0.0523 .
## Year2004          0.07618      0.03730      2.04      0.0412 *
## Year2005          0.04483      0.03613      1.24      0.2147
## Year2006          0.06869      0.03623      1.90      0.0580 .
## Year2007          0.06662      0.03565      1.87      0.0617 .
## Year2008          0.15288      0.03672      4.16     3.2e-05 ***
## Year2009          0.08427      0.03611      2.33      0.0196 *
## Year2010          0.07367      0.03674      2.01      0.0450 *
## Year2011          0.12169      0.03778      3.22      0.0013 **
## Year2012          0.18695      0.03740      5.00     5.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.0314, Adjusted R-squared:  0.0294
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 894 weights are ~= 1. The remaining 9981 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0067 0.8700 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.20e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.437 1          1.199
## LastAuthorFemale 1.433 1          1.197
## Year              1.017 16          1.001

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1786 0001438979 3.878 1998      2002      2      2.79
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34957 -0.40618  0.00825  0.41329  2.78996
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03624    0.02856   36.29 < 2e-16 ***
## FirstAuthorFemale1 0.01381    0.01769    0.78  0.43479
## LastAuthorFemale1 0.08366    0.01782    4.69  2.7e-06 ***
## Year1997        0.05990    0.03861    1.55  0.12089
## Year1998        0.05180    0.03946    1.31  0.18935
## Year1999        0.11071    0.03884    2.85  0.00437 **
## Year2000        0.05886    0.03689    1.60  0.11060
## Year2001        0.00618    0.04208    0.15  0.88327
## Year2002        0.04521    0.03615    1.25  0.21114
## Year2003        0.09351    0.03861    2.42  0.01546 *
## Year2004        0.09530    0.03759    2.54  0.01126 *
## Year2005        0.07033    0.03627    1.94  0.05248 .
```

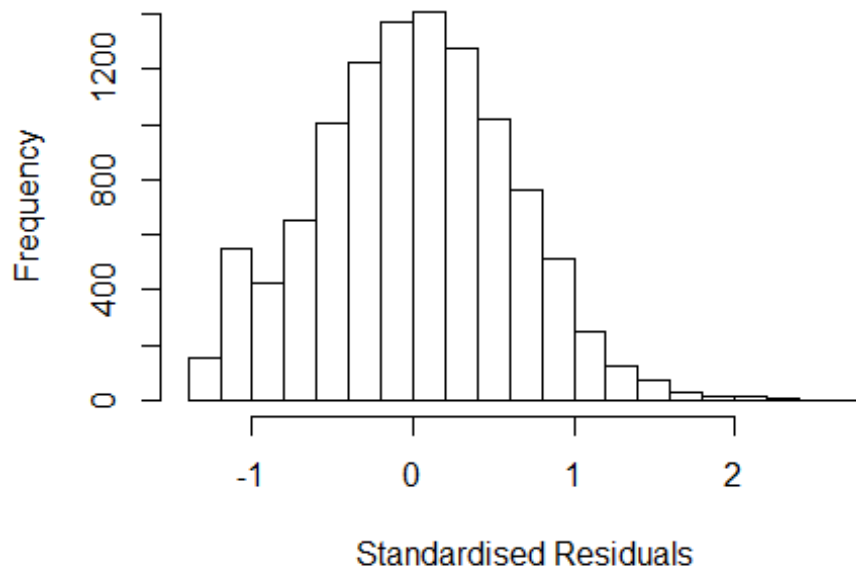


```

## Year2006          0.09016      0.03651      2.47  0.01355 *
## Year2007          0.09142      0.03589      2.55  0.01087 *
## Year2008          0.18081      0.03700      4.89  1.0e-06 ***
## Year2009          0.10355      0.03640      2.84  0.00446 **
## Year2010          0.10594      0.03693      2.87  0.00413 **
## Year2011          0.14834      0.03834      3.87  0.00011 ***
## Year2012          0.21585      0.03778      5.71  1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.0121, Adjusted R-squared:  0.0105
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 915 weights are ~= 1. The remaining 9960 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0023 0.8660 0.9510 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.20e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1      1.006
## Year              1.012 16      1.000

```

## Residuals from first author



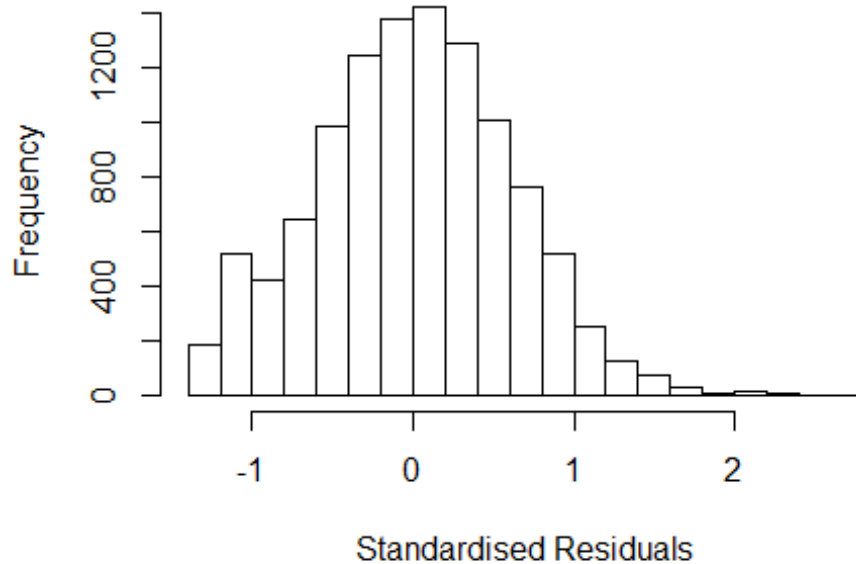
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1786 0001438979 3.878 1998      2002      2      2.79
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32132 -0.41126  0.00545  0.41670  2.78321
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04158    0.02855   36.48 < 2e-16 ***
## FirstAuthorFemale1 0.06034    0.01481    4.07 4.6e-05 ***
## Year1997        0.06002    0.03862    1.55  0.1202
## Year1998        0.05321    0.03957    1.34  0.1787
## Year1999        0.11156    0.03889    2.87  0.0041 **
## Year2000        0.05926    0.03692    1.61  0.1085
## Year2001        0.00864    0.04213    0.21  0.8374
## Year2002        0.04897    0.03621    1.35  0.1762
## Year2003        0.09410    0.03864    2.44  0.0149 *
## Year2004        0.09752    0.03761    2.59  0.0095 **
## Year2005        0.07403    0.03631    2.04  0.0415 *
## Year2006        0.09068    0.03653    2.48  0.0131 *
```

```

## Year2007          0.09299    0.03594    2.59    0.0097 **
## Year2008          0.18177    0.03709    4.90    9.7e-07 ***
## Year2009          0.10603    0.03648    2.91    0.0037 **
## Year2010          0.10881    0.03697    2.94    0.0033 **
## Year2011          0.15142    0.03838    3.95    8.0e-05 ***
## Year2012          0.21940    0.03782    5.80    6.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.611
## Multiple R-squared:  0.0101, Adjusted R-squared:  0.0085
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 887 weights are ~= 1. The remaining 9988 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0032 0.8690 0.9510 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.20e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.009 1          1.004
## Year            1.009 16          1.000

```

## Residuals from last author



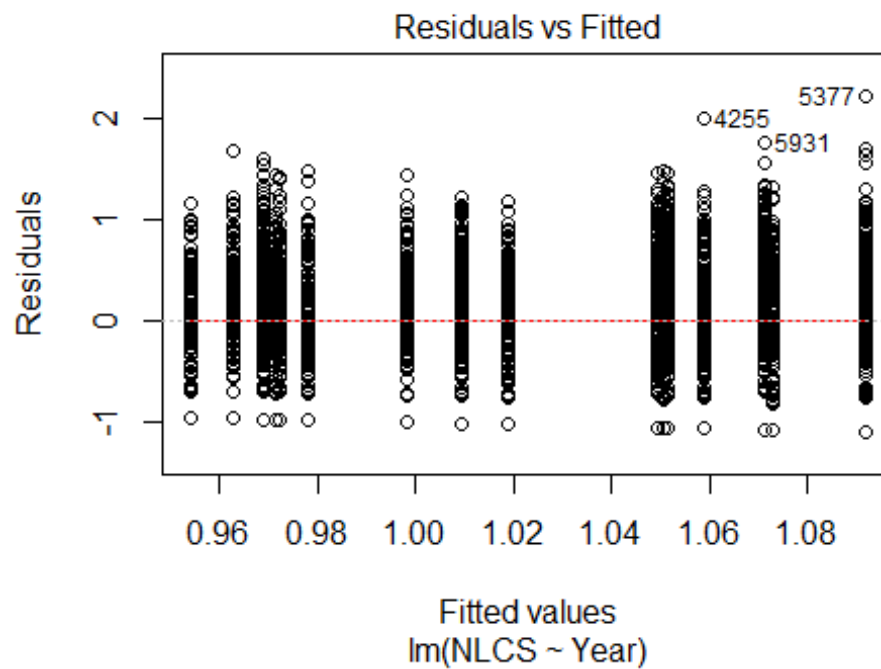
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1786 0001438979 3.878 1998      2002      2      2.79
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34501 -0.40607  0.00832  0.41415  2.78904
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03687    0.02855   36.32  < 2e-16 ***
## LastAuthorFemale1 0.09144    0.01495    6.12  9.8e-10 ***
## Year1997        0.06033    0.03861    1.56  0.1181
## Year1998        0.05209    0.03945    1.32  0.1868
## Year1999        0.11094    0.03883    2.86  0.0043 **
## Year2000        0.05944    0.03688    1.61  0.1070
## Year2001        0.00611    0.04208    0.15  0.8845
## Year2002        0.04596    0.03613    1.27  0.2034
## Year2003        0.09385    0.03862    2.43  0.0151 *
## Year2004        0.09600    0.03758    2.55  0.0107 *
## Year2005        0.07069    0.03626    1.95  0.0513 .
## Year2006        0.09138    0.03648    2.51  0.0123 *
```

```

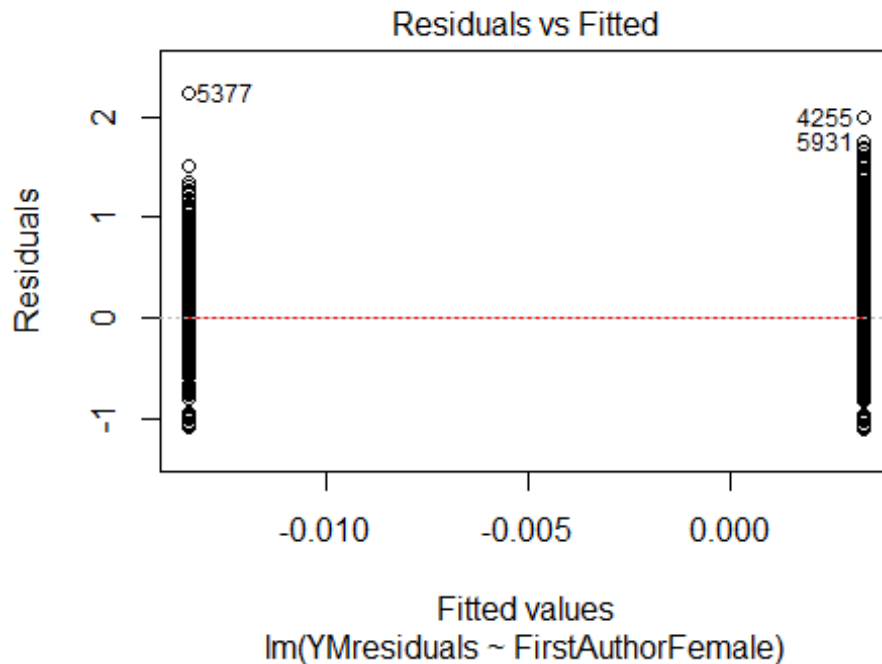
## Year2007      0.09205      0.03586      2.57      0.0103 *
## Year2008      0.18189      0.03696      4.92      8.7e-07 ***
## Year2009      0.10457      0.03637      2.88      0.0040 **
## Year2010      0.10681      0.03691      2.89      0.0038 **
## Year2011      0.14905      0.03830      3.89      0.0001 ***
## Year2012      0.21669      0.03775      5.74      9.7e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.0121, Adjusted R-squared:  0.0105
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 917 weights are ~= 1. The remaining 9958 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0023 0.8660 0.9500 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.20e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 10875"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2003"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 187 209 215 238 245 235 253 155 255 261 304 321 350 386 378
## 2011 2012
## 359 398
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 148 164 174 193 192 157 213 131 205 215 256 257 275 291 296
## 2011 2012

```

```
## 282 298
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 144 154 167 185 181 149 203 123 190 202 235 233 252 266 267
## 2011 2012
## 257 266
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```

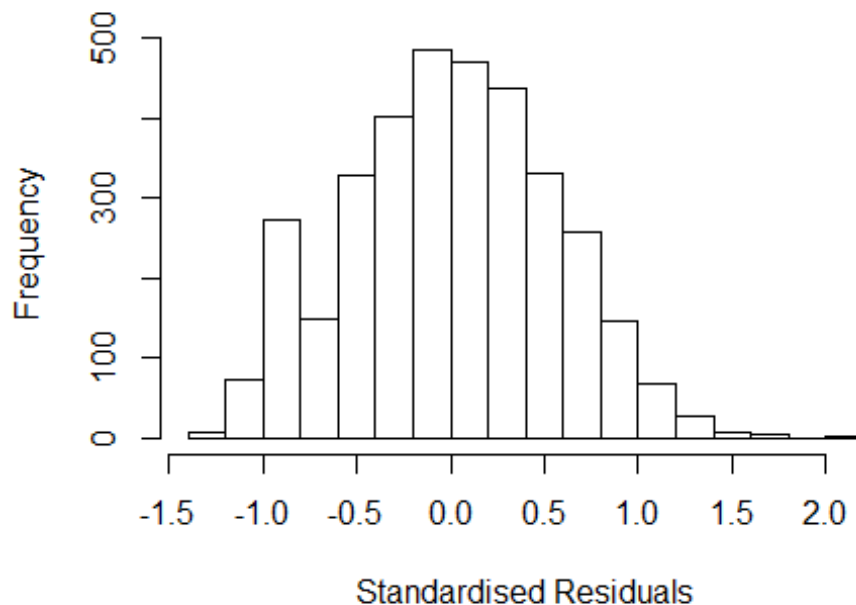


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.68, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 1.68398550119266"
## [1] "Male first author team size 2018 geometric mean: 1.64028413713912"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5900, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.59460808267637"
## [1] "Male last author team size 2018 geometric mean: 1.66238086680143"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5500, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.298 1          1.139
## LastAuthorFemale  1.272 1          1.128
## UniqueAuthors     1.104 4          1.012
## Year              1.108 16         1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2985 -0.3795  0.0046  0.3839  2.1935
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.88639    0.05287   16.77 < 2e-16 ***
## FirstAuthorFemale1 -0.01946    0.02805   -0.69  0.488
## LastAuthorFemale1 -0.00600    0.03008   -0.20  0.842
## UniqueAuthors2    0.18592    0.02245    8.28 < 2e-16 ***
## UniqueAuthors3    0.34840    0.02868   12.15 < 2e-16 ***
## UniqueAuthors4    0.27990    0.06781    4.13 3.8e-05 ***
## UniqueAuthors5    0.37509    0.15673    2.39  0.017 *
## Year1997          0.00440    0.06958    0.06  0.950
## Year1998         -0.00671    0.07020   -0.10  0.924
## Year1999          0.11017    0.06799    1.62  0.105
```

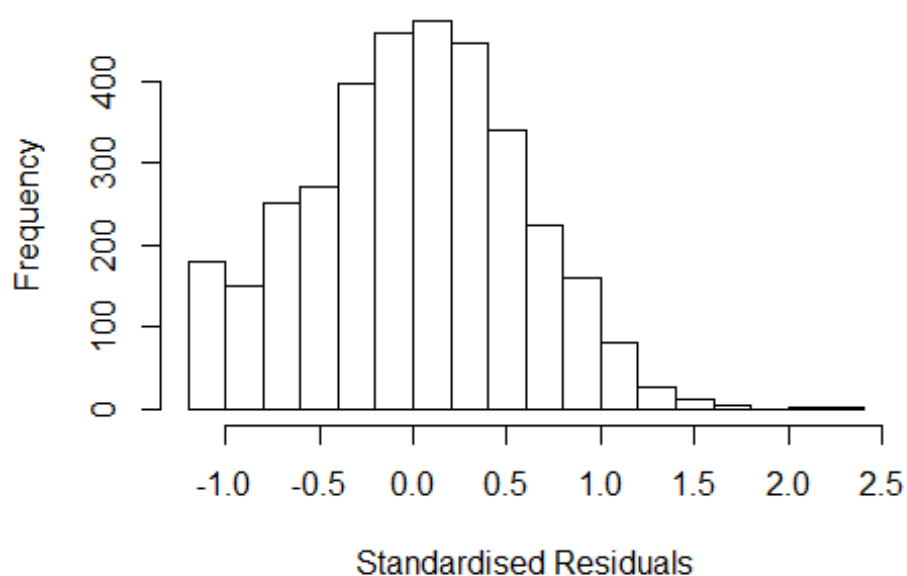


```

## Year2000      -0.02740    0.06766   -0.40    0.686
## Year2001      -0.00356    0.06980   -0.05    0.959
## Year2002       0.00694    0.06618    0.10    0.917
## Year2003      -0.03046    0.06985   -0.44    0.663
## Year2004      -0.02964    0.06648   -0.45    0.656
## Year2005       0.03065    0.06547    0.47    0.640
## Year2006      -0.04368    0.06294   -0.69    0.488
## Year2007      -0.00494    0.06374   -0.08    0.938
## Year2008       0.05305    0.06671    0.80    0.427
## Year2009       0.04661    0.06292    0.74    0.459
## Year2010       0.00938    0.06521    0.14    0.886
## Year2011       0.06368    0.06577    0.97    0.333
## Year2012       0.05901    0.06548    0.90    0.368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.0538, Adjusted R-squared:  0.0478
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 281 weights are ~= 1. The remaining 3193 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.107  0.869  0.950  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.274 1      1.129
## LastAuthorFemale  1.281 1      1.132
## Year              1.030 16      1.001

```

## Residuals from first and last author



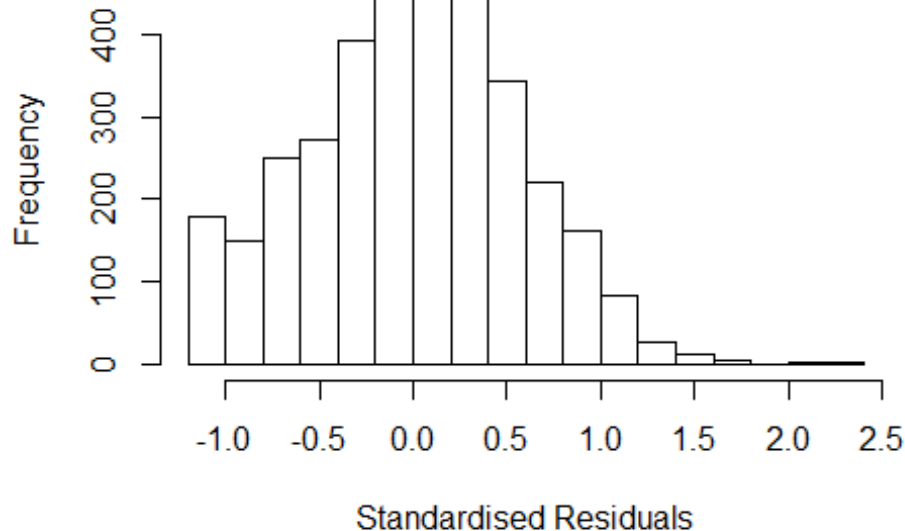
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0899 -0.3877 0.0138 0.3918 2.2519
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94710 0.05297 17.88 <2e-16 ***
## FirstAuthorFemale1 -0.00366 0.02884 -0.13 0.899
## LastAuthorFemale1 -0.00740 0.03112 -0.24 0.812
## Year1997 0.02458 0.07081 0.35 0.729
## Year1998 0.02459 0.07135 0.34 0.730
## Year1999 0.14277 0.06893 2.07 0.038 *
## Year2000 0.00860 0.06778 0.13 0.899
## Year2001 0.04315 0.07155 0.60 0.546
## Year2002 0.07428 0.06578 1.13 0.259
## Year2003 0.01073 0.07154 0.15 0.881
## Year2004 0.01987 0.06688 0.30 0.766
## Year2005 0.10008 0.06578 1.52 0.128
```

```

## Year2006          0.00437    0.06378    0.07    0.945
## Year2007          0.05983    0.06493    0.92    0.357
## Year2008          0.11170    0.06734    1.66    0.097 .
## Year2009          0.10302    0.06342    1.62    0.104
## Year2010          0.08502    0.06557    1.30    0.195
## Year2011          0.11463    0.06698    1.71    0.087 .
## Year2012          0.11927    0.06679    1.79    0.074 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.00618,    Adjusted R-squared:  0.00101
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 284 weights are ~= 1. The remaining 3190 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0944 0.8710 0.9510 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.016 1      1.008
## Year              1.016 16      1.000

```

## Residuals from first author



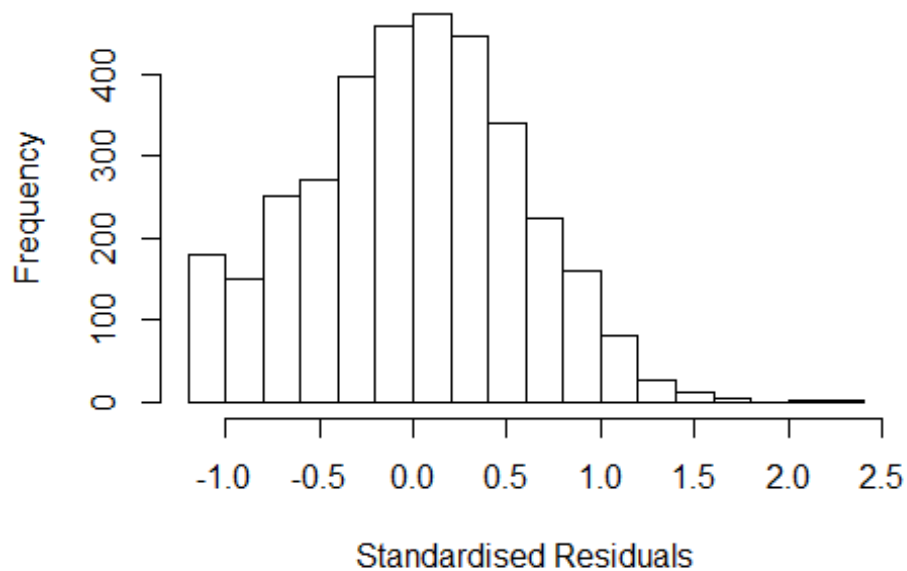
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0892 -0.3879 0.0145 0.3912 2.2562
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94670 0.05295 17.88 <2e-16 ***
## FirstAuthorFemale1 -0.00735 0.02588 -0.28 0.776
## Year1997 0.02457 0.07081 0.35 0.729
## Year1998 0.02441 0.07132 0.34 0.732
## Year1999 0.14246 0.06890 2.07 0.039 *
## Year2000 0.00855 0.06777 0.13 0.900
## Year2001 0.04305 0.07153 0.60 0.547
## Year2002 0.07391 0.06573 1.12 0.261
## Year2003 0.01078 0.07152 0.15 0.880
## Year2004 0.01951 0.06684 0.29 0.770
## Year2005 0.09960 0.06572 1.52 0.130
## Year2006 0.00422 0.06377 0.07 0.947
```

```

## Year2007          0.05972      0.06492      0.92      0.358
## Year2008          0.11139      0.06730      1.65      0.098 .
## Year2009          0.10265      0.06336      1.62      0.105
## Year2010          0.08495      0.06556      1.30      0.195
## Year2011          0.11448      0.06696      1.71      0.087 .
## Year2012          0.11904      0.06673      1.78      0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.00616,    Adjusted R-squared:  0.00127
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 285 weights are ~= 1. The remaining 3189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0933 0.8710 0.9510 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.02 1      1.010
## Year            1.02 16      1.001

```

## Residuals from last author



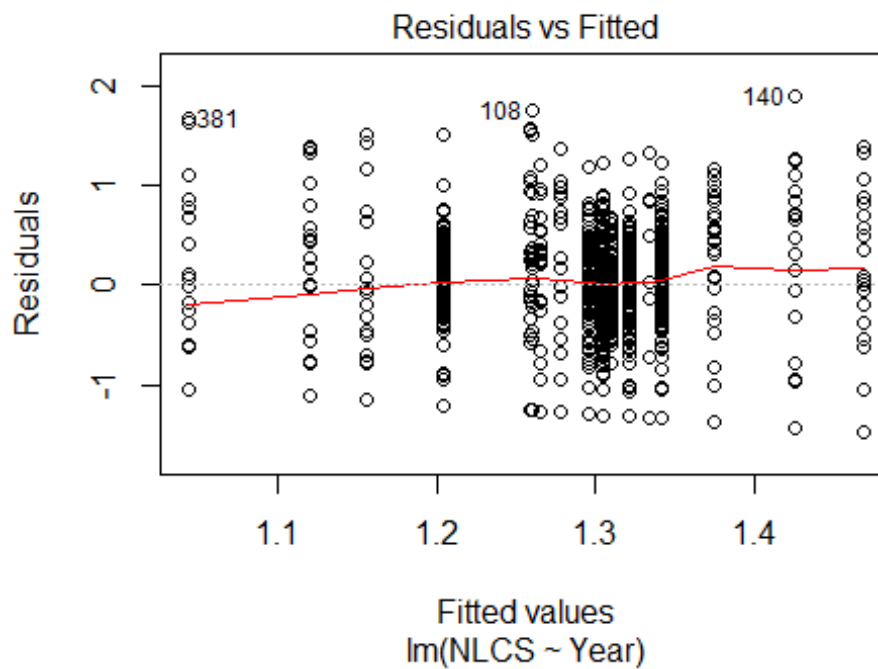
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0895 -0.3876 0.0146 0.3901 2.2486
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94694 0.05295 17.89 <2e-16 ***
## LastAuthorFemale1 -0.00935 0.02781 -0.34 0.737
## Year1997 0.02423 0.07075 0.34 0.732
## Year1998 0.02458 0.07136 0.34 0.731
## Year1999 0.14260 0.06895 2.07 0.039 *
## Year2000 0.00833 0.06781 0.12 0.902
## Year2001 0.04303 0.07155 0.60 0.548
## Year2002 0.07397 0.06573 1.13 0.261
## Year2003 0.01046 0.07153 0.15 0.884
## Year2004 0.01978 0.06688 0.30 0.767
## Year2005 0.10002 0.06578 1.52 0.128
## Year2006 0.00406 0.06374 0.06 0.949
```

```

## Year2007      0.05963      0.06494      0.92      0.359
## Year2008      0.11144      0.06736      1.65      0.098 .
## Year2009      0.10278      0.06342      1.62      0.105
## Year2010      0.08473      0.06551      1.29      0.196
## Year2011      0.11445      0.06697      1.71      0.088 .
## Year2012      0.11895      0.06675      1.78      0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.00618,    Adjusted R-squared:  0.00129
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 281 weights are ~= 1. The remaining 3193 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0958 0.8710 0.9510 0.9100 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3474"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2100"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   73   51   71   50   56   42   46   42   47   41   56   157  225  152  201
## 2011 2012
##  156  259
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   26   21   22   20   21    9   20   23   33   24   32  106  164   99  128
## 2011 2012

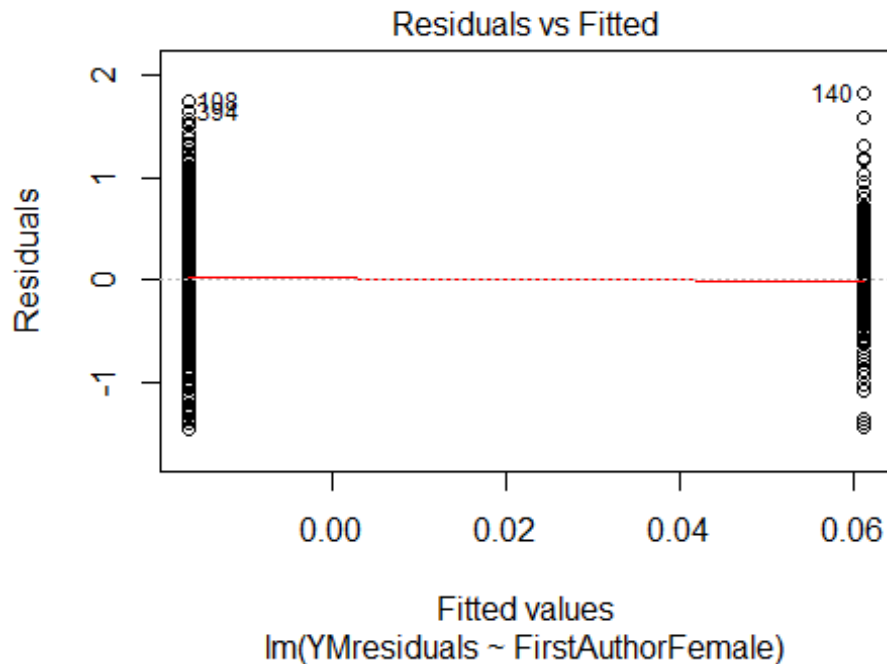
```

```
## 105 168
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 21 21 20 20 8 20 23 30 22 31 96 153 90 117
## 2011 2012
## 93 144
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 230, df = 16, p-value <2e-16
```



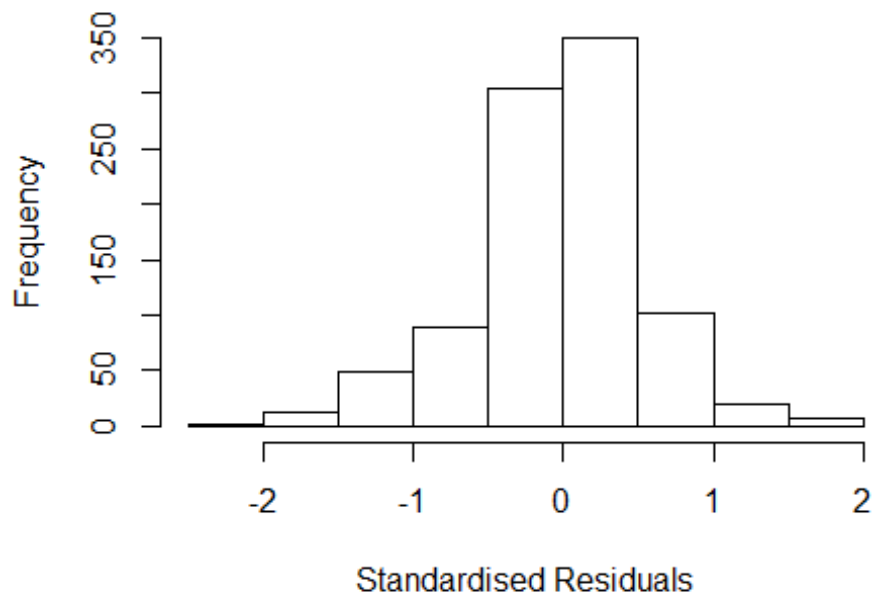
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.5, df = 1, p-value = 0.004
```





```
## [1] "Female first author team size 2018 geometric mean: 3.24508963960429"
## [1] "Male first author team size 2018 geometric mean: 2.83613151842995"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.30639565054252"
## [1] "Male last author team size 2018 geometric mean: 2.85812046218049"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.229 1          1.108
## LastAuthorFemale  1.330 1          1.153
## UniqueAuthors     1.773 4          1.074
## Year               2.618 16         1.031
```

## Residuals from first and last author and team size



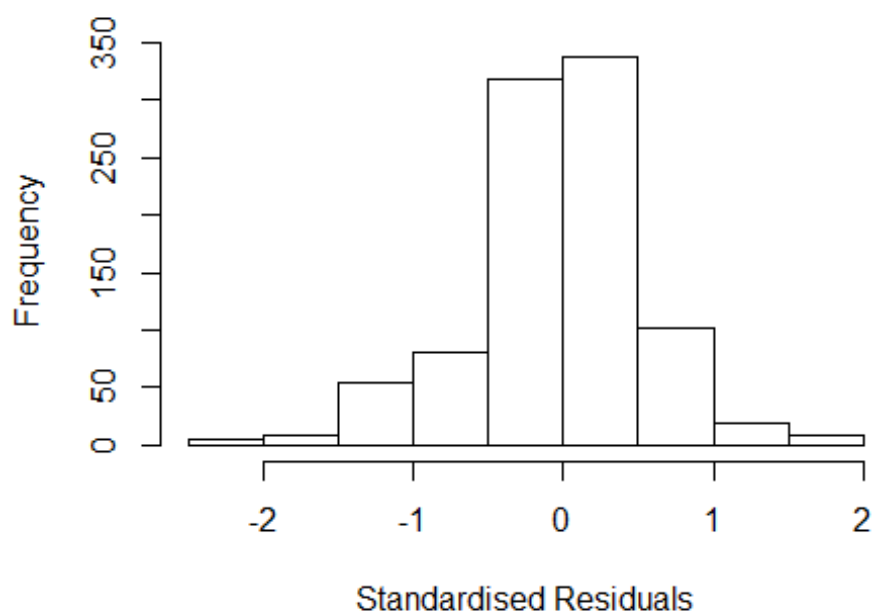
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.0289 -0.3029 0.0167 0.2992 1.9102
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3969 0.1865 7.49 1.6e-13 ***
## FirstAuthorFemale1 0.0345 0.0395 0.87 0.3827
## LastAuthorFemale1 0.0509 0.0543 0.94 0.3490
## UniqueAuthors2 0.1669 0.0589 2.83 0.0047 **
## UniqueAuthors3 0.0333 0.0609 0.55 0.5852
## UniqueAuthors4 0.1332 0.0623 2.14 0.0328 *
## UniqueAuthors5 0.1244 0.0559 2.23 0.0263 *
## Year1997 -0.2863 0.2581 -1.11 0.2676
## Year1998 0.4651 0.4274 1.09 0.2768
## Year1999 -0.4349 0.2726 -1.60 0.1110
```

```

## Year2000          -0.1856      0.3632   -0.51   0.6094
## Year2001           0.2657      1.1460    0.23   0.8167
## Year2002          -0.6291      0.3445   -1.83   0.0682 .
## Year2003          -0.0495      0.4278   -0.12   0.9079
## Year2004           0.2642      0.2832    0.93   0.3511
## Year2005          -0.3827      0.3262   -1.17   0.2411
## Year2006          -0.0610      0.2328   -0.26   0.7934
## Year2007          -0.2621      0.1885   -1.39   0.1647
## Year2008          -0.1834      0.1853   -0.99   0.3225
## Year2009          -0.1937      0.1860   -1.04   0.2980
## Year2010          -0.2010      0.1860   -1.08   0.2803
## Year2011          -0.2210      0.1872   -1.18   0.2379
## Year2012          -0.1493      0.1859   -0.80   0.4221
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.1, Adjusted R-squared:  0.0785
## Convergence in 46 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 853 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0132 0.8490 0.9520 0.8690 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.07e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.162 1 1.078
## LastAuthorFemale 1.263 1 1.124
## Year 1.377 16 1.010

```

## Residuals from first and last author



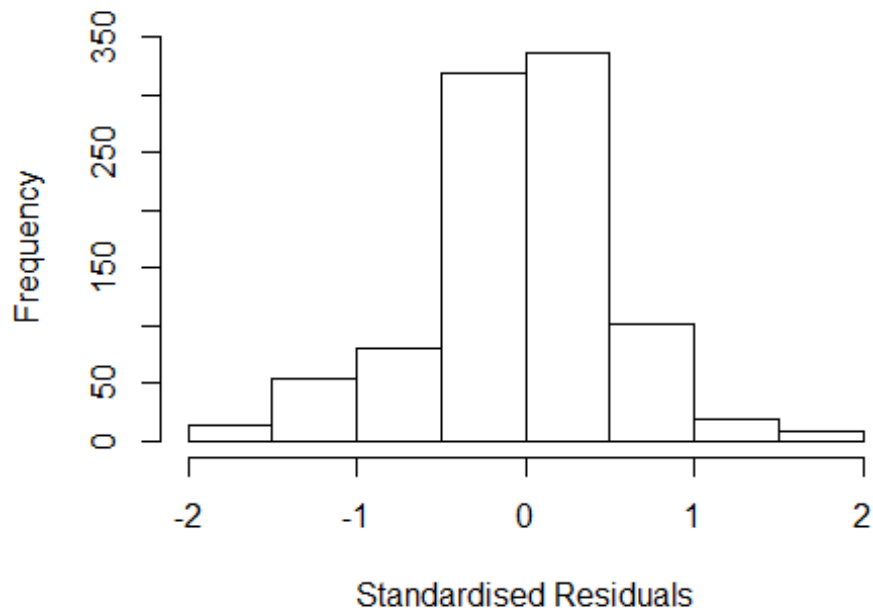
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.00224 -0.31315 -0.00123 0.30568 1.86328
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4874 0.1827 8.14 1.3e-15 ***
## FirstAuthorFemale1 0.0477 0.0379 1.26 0.208
## LastAuthorFemale1 0.0467 0.0523 0.89 0.372
## Year1997 -0.2570 0.2648 -0.97 0.332
## Year1998 0.5148 0.3383 1.52 0.128
## Year1999 -0.4944 0.2795 -1.77 0.077 .
## Year2000 -0.2450 0.3936 -0.62 0.534
## Year2001 0.3185 0.7104 0.45 0.654
## Year2002 -0.6727 0.3483 -1.93 0.054 .
## Year2003 -0.1088 0.4080 -0.27 0.790
## Year2004 0.2486 0.2748 0.90 0.366
## Year2005 -0.4255 0.3225 -1.32 0.187
```

```

## Year2006          -0.0857      0.2333   -0.37    0.714
## Year2007          -0.2466      0.1887   -1.31    0.192
## Year2008          -0.1925      0.1858   -1.04    0.301
## Year2009          -0.1943      0.1865   -1.04    0.298
## Year2010          -0.1959      0.1867   -1.05    0.294
## Year2011          -0.2172      0.1885   -1.15    0.249
## Year2012          -0.1559      0.1866   -0.84    0.404
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.46
## Multiple R-squared:  0.0908, Adjusted R-squared:  0.0729
## Convergence in 45 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 853 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0184 0.8480 0.9500 0.8670 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.07e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.132 1      1.064
## Year              1.132 16      1.004

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.99460 -0.31725 -0.00391 0.30777 1.85135
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4879 0.1835 8.11 1.6e-15 ***
## FirstAuthorFemale1 0.0582 0.0378 1.54 0.124
## Year1997 -0.2552 0.2657 -0.96 0.337
## Year1998 0.5067 0.3486 1.45 0.146
## Year1999 -0.4778 0.2782 -1.72 0.086 .
## Year2000 -0.2500 0.3838 -0.65 0.515
## Year2001 0.2813 0.7559 0.37 0.710
## Year2002 -0.6613 0.3437 -1.92 0.055 .
## Year2003 -0.1113 0.4028 -0.28 0.782
## Year2004 0.2372 0.2775 0.85 0.393
## Year2005 -0.4175 0.3238 -1.29 0.198
## Year2006 -0.0857 0.2345 -0.37 0.715
```

```

## Year2007          -0.2405      0.1895   -1.27    0.205
## Year2008          -0.1907      0.1867   -1.02    0.307
## Year2009          -0.1897      0.1872   -1.01    0.311
## Year2010          -0.1932      0.1875   -1.03    0.303
## Year2011          -0.2119      0.1892   -1.12    0.263
## Year2012          -0.1516      0.1874   -0.81    0.419
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.463
## Multiple R-squared:  0.0862, Adjusted R-squared:  0.0693
## Convergence in 36 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0234 0.8470 0.9490 0.8680 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.07e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.219 1      1.104
## Year      1.219 16      1.006
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -2.00363 -0.31905  0.00771  0.30699  1.86222

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4931    0.1821   8.20 8.2e-16 ***
## LastAuthorFemale1  0.0615    0.0513   1.20  0.231
## Year1997         -0.2546    0.2655  -0.96  0.338
## Year1998          0.5106    0.3477   1.47  0.142
## Year1999         -0.4914    0.2782  -1.77  0.078 .
## Year2000         -0.2437    0.3875  -0.63  0.530
## Year2001          0.3408    0.6563   0.52  0.604
## Year2002         -0.6773    0.3462  -1.96  0.051 .
## Year2003         -0.1100    0.3999  -0.27  0.783
## Year2004          0.2550    0.2733   0.93  0.351
## Year2005         -0.4136    0.3193  -1.30  0.196
## Year2006         -0.0860    0.2332  -0.37  0.712
## Year2007         -0.2462    0.1882  -1.31  0.191
## Year2008         -0.1885    0.1852  -1.02  0.309
## Year2009         -0.1888    0.1859  -1.02  0.310
## Year2010         -0.1888    0.1859  -1.02  0.310
## Year2011         -0.2156    0.1879  -1.15  0.252
## Year2012         -0.1532    0.1860  -0.82  0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0886, Adjusted R-squared:  0.0717
## Convergence in 32 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0196 0.8480 0.9500 0.8670 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.07e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 935"
## [1] ""

```



```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2101"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      8   10    4    8   15    9   13    3   18   18   17   17   15    8    8
## 2011 2012
##     10   19
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      1    2    1    2    2    1    6    2   12   15   12    9    3    8    8
## 2011 2012
##      8   17
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      1    2    1    2    2    1    6    1   11   15   11    9    3    7    8
## 2011 2012
##      7   16
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.22446470709902"
## [1] "Male first author team size 2018 geometric mean: 2.41234712184771"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

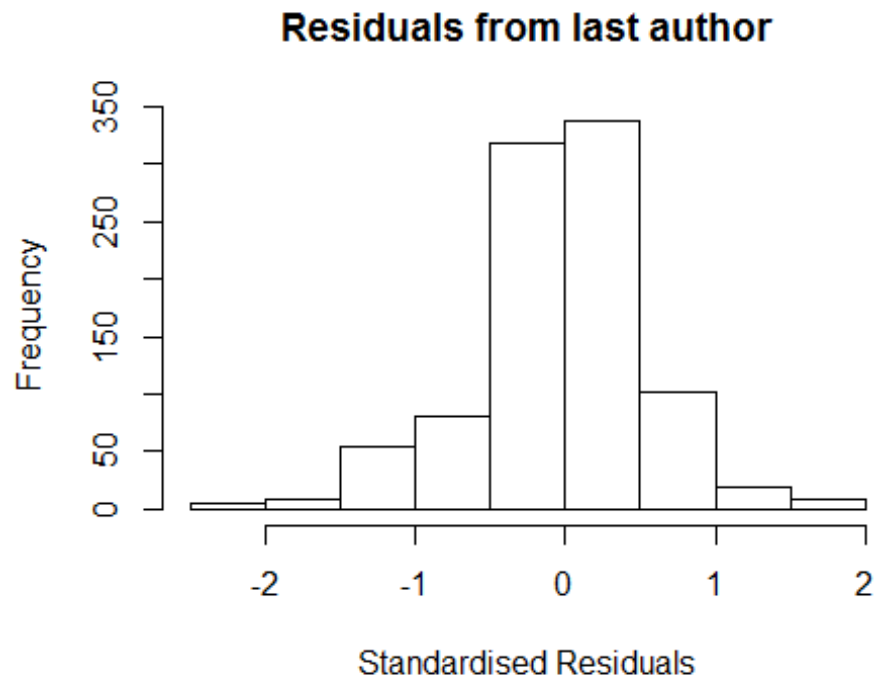
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.75892417638112"
## [1] "Male last author team size 2018 geometric mean: 2.60211822180536"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 53, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

```

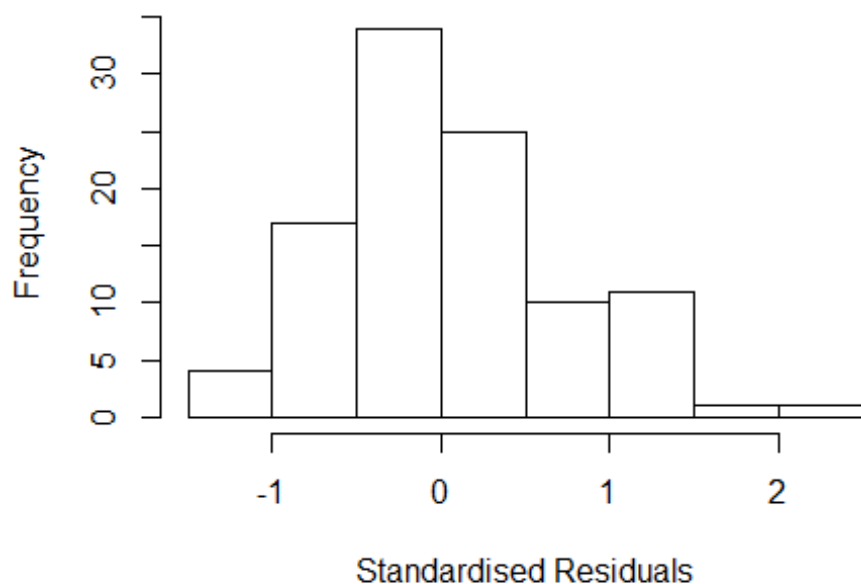
```
## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate
```



```
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
## FirstAuthorFemale	-4.417e+14	1	NaN
## LastAuthorFemale	-5.365e+12	1	NaN
## Year	-4.049e+28	16	NaN

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31e+00 -3.75e-01 5.55e-16 2.77e-01 2.02e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.395 0.416 -0.95 0.3446
## FirstAuthorFemale1 0.600 0.449 1.34 0.1849
## LastAuthorFemale1 -0.205 0.376 -0.54 0.5873
## Year1997 0.697 0.472 1.48 0.1434
## Year1998 1.936 0.416 4.66 1.2e-05 ***
## Year1999 0.395 0.416 0.95 0.3446
## Year2000 1.205 0.960 1.26 0.2128
## Year2001 1.370 0.449 3.05 0.0030 **
## Year2002 0.509 0.421 1.21 0.2299
## Year2003 1.844 0.416 4.43 2.8e-05 ***
## Year2004 1.114 0.390 2.85 0.0054 **
## Year2005 1.309 0.544 2.41 0.0182 *
```

```

## Year2006          0.768      0.425      1.81      0.0740 .
## Year2007          1.204      0.475      2.54      0.0131 *
## Year2008          1.307      0.746      1.75      0.0833 .
## Year2009          1.433      0.462      3.10      0.0026 **
## Year2010          1.073      0.377      2.85      0.0055 **
## Year2011          0.990      0.437      2.27      0.0259 *
## Year2012          1.089      0.378      2.88      0.0051 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.22, Adjusted R-squared:  0.053
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 87 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
## 0.257 0.817 0.939 0.885 0.990 0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.71e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -4.946e+16 1      NaN
## Year -4.946e+16 16      NaN
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max

```

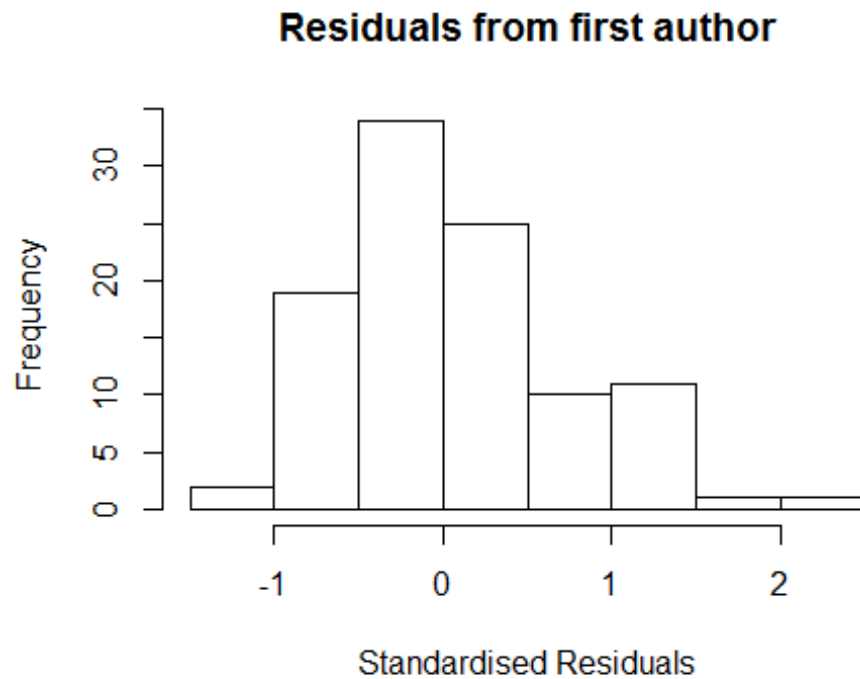
```

## -1.386 -0.375 0.000 0.279 2.035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.472     0.412   -1.15  0.25454
## FirstAuthorFemale1  0.472     0.412    1.15  0.25454
## Year1997       0.774     0.469    1.65  0.10214
## Year1998       2.013     0.412    4.89  4.7e-06 ***
## Year1999       0.472     0.412    1.15  0.25454
## Year2000       1.282     0.980    1.31  0.19428
## Year2001       1.242     0.412    3.02  0.00337 **
## Year2002       0.586     0.417    1.41  0.16341
## Year2003       1.921     0.412    4.67  1.1e-05 ***
## Year2004       1.176     0.396    2.97  0.00385 **
## Year2005       1.386     0.549    2.52  0.01344 *
## Year2006       0.843     0.423    1.99  0.04954 *
## Year2007       1.280     0.500    2.56  0.01228 *
## Year2008       1.408     0.684    2.06  0.04245 *
## Year2009       1.470     0.463    3.17  0.00209 **
## Year2010       1.152     0.320    3.60  0.00053 ***
## Year2011       1.031     0.457    2.26  0.02652 *
## Year2012       1.165     0.371    3.14  0.00233 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.218, Adjusted R-squared:  0.0618
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 88 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.241  0.822  0.946  0.884  0.989  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          9.71e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"

```

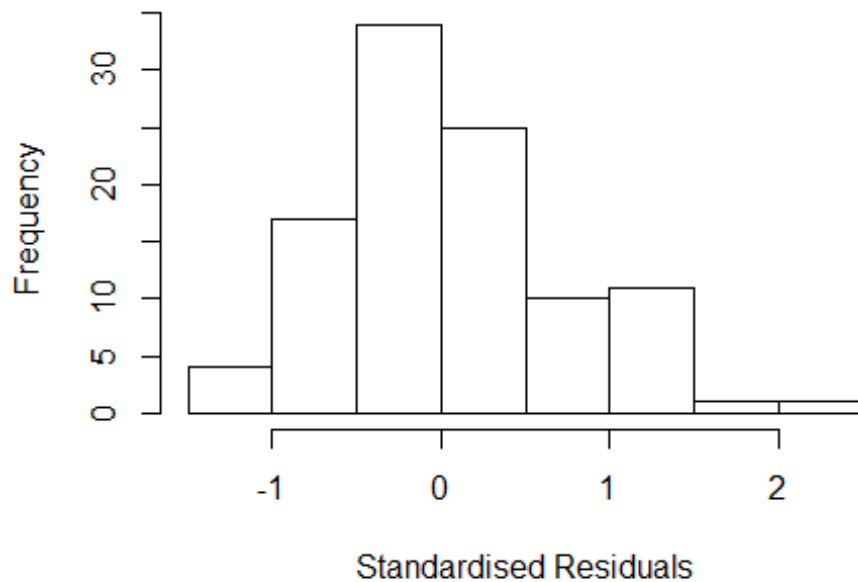
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale  NaN  1          NaN
## Year              NaN 16          NaN
```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.09e+00 -4.11e-01 -2.22e-16 3.99e-01 1.97e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.170 0.297 -0.57 0.5695
## LastAuthorFemale1 0.170 0.297 0.57 0.5695
## Year1997 0.472 0.371 1.27 0.2075
## Year1998 1.711 0.297 5.76 1.3e-07 ***
## Year1999 0.170 0.297 0.57 0.5695
## Year2000 0.980 0.895 1.09 0.2766
## Year2001 0.770 0.000 Inf < 2e-16 ***
## Year2002 0.283 0.304 0.93 0.3545
## Year2003 1.619 0.297 5.45 4.8e-07 ***
## Year2004 0.994 0.407 2.44 0.0166 *
## Year2005 1.090 0.430 2.54 0.0130 *
## Year2006 0.551 0.347 1.59 0.1159
```

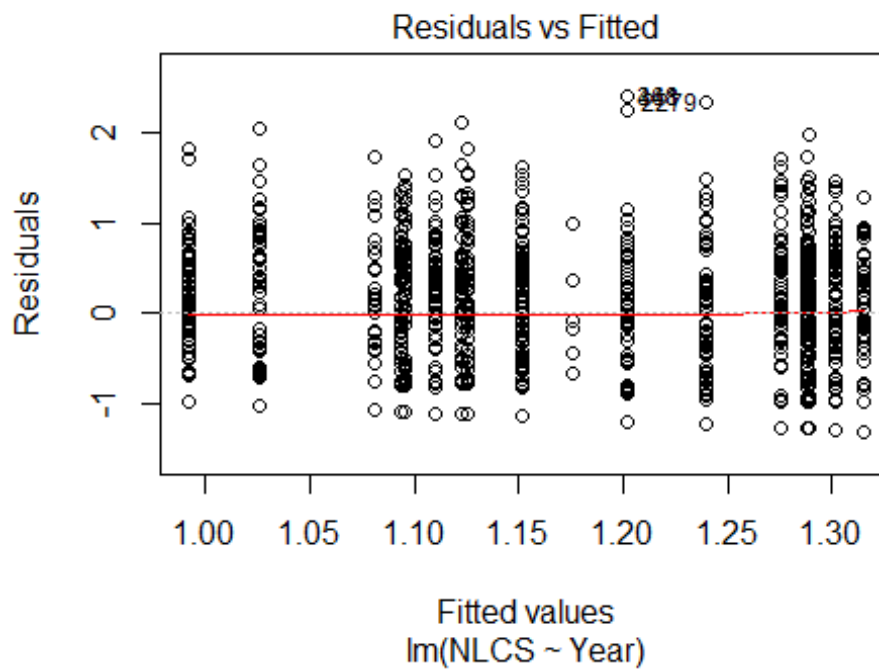
```

## Year2007          0.959      0.373      2.57      0.0118 *
## Year2008          1.232      0.460      2.68      0.0088 **
## Year2009          1.130      0.337      3.35      0.0012 **
## Year2010          1.003      0.391      2.56      0.0121 *
## Year2011          0.703      0.327      2.15      0.0342 *
## Year2012          0.946      0.298      3.17      0.0021 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.631
## Multiple R-squared:  0.182, Adjusted R-squared:  0.018
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 88 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.311  0.846  0.938  0.889  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.71e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 103"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2102"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  295  274  266   26  320  228  203  168  185  163  223  245  223  285  253
## 2011 2012
##  259  319
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   63   62   66    6   62   34   62   53   61   56   82   94   75  118   97
## 2011 2012

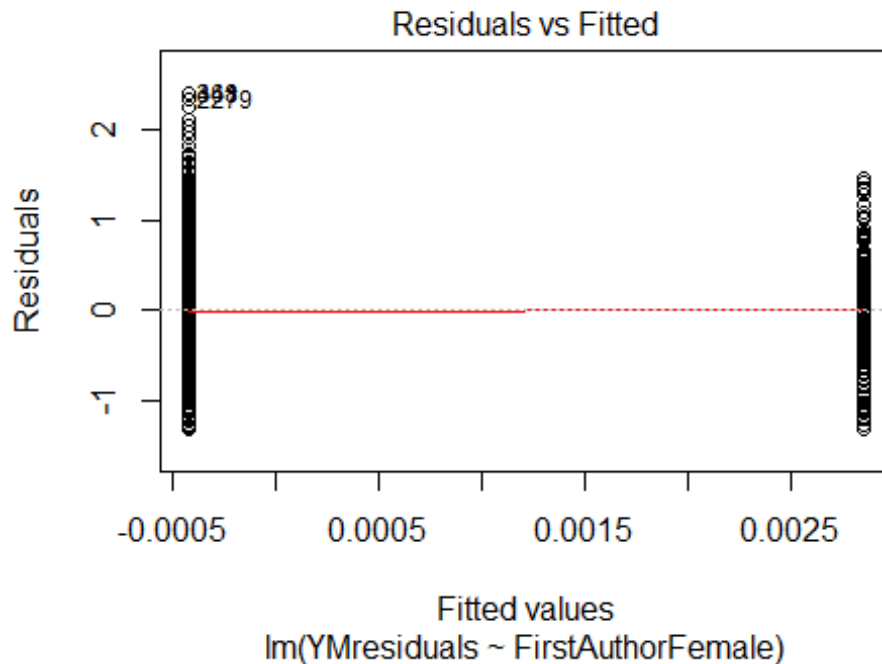
```



```
##      88   132
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   54   54   57    6   54   32   55   44   55   51   72   76   65   99   78
## 2011 2012
##   77   112
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```

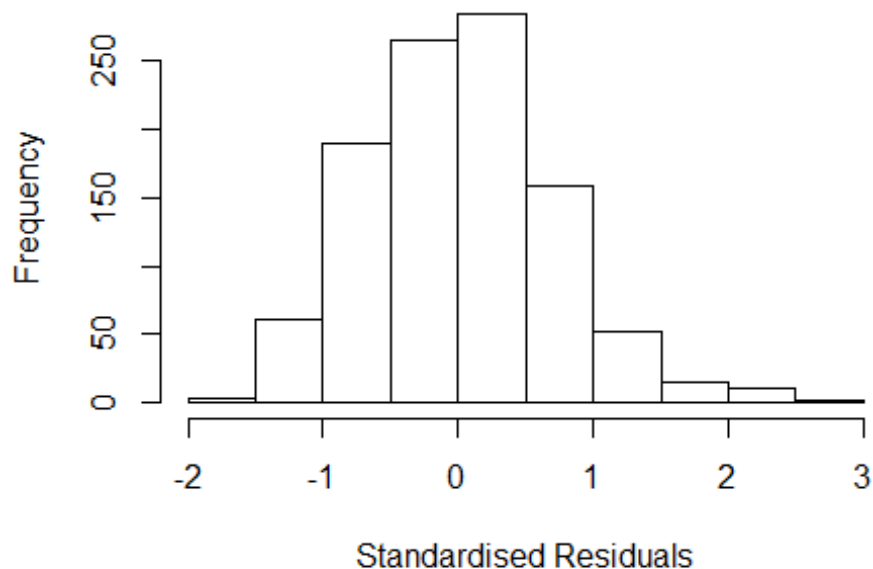


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.6, df = 1, p-value = 0.06
```



```
## [1] "Female first author team size 2018 geometric mean: 2.80862047363036"
## [1] "Male first author team size 2018 geometric mean: 3.22954966856366"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.10109212965504"
## [1] "Male last author team size 2018 geometric mean: 3.31269594305551"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 2e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.258 1          1.122
## LastAuthorFemale  1.253 1          1.120
## UniqueAuthors    1.409 4          1.044
## Year              1.506 16         1.013
```

## Residuals from first and last author and team size



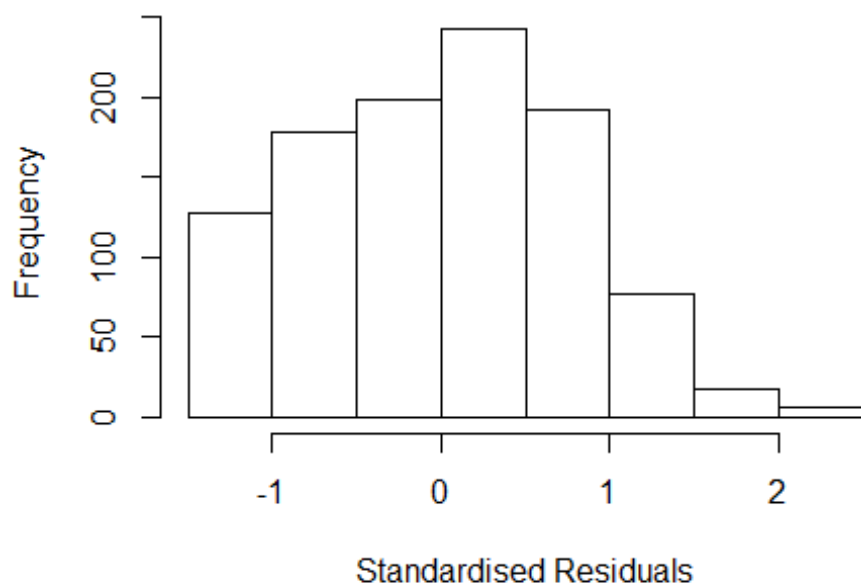
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 366  0031200253 3.451 1997    2102     2    2.787
## 2060 0038180782 3.269 2003    2102     2    2.660
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6262 -0.4788  0.0163  0.4456  2.7867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.63512    0.11001   5.77   1e-08 ***
## FirstAuthorFemale1 -0.09222    0.06908  -1.33   0.182
## LastAuthorFemale1  0.06725    0.09364   0.72   0.473
## UniqueAuthors2    0.64857    0.06872   9.44 <2e-16 ***
## UniqueAuthors3    0.83809    0.06180  13.56 <2e-16 ***
## UniqueAuthors4    0.83752    0.07653  10.94 <2e-16 ***
## UniqueAuthors5    0.85832    0.07004  12.25 <2e-16 ***
## Year1997          0.02913    0.15696   0.19   0.853
## Year1998         -0.00755    0.14351  -0.05   0.958
```

```

## Year1999      0.03449      0.25385      0.14      0.892
## Year2000      0.04806      0.15429      0.31      0.755
## Year2001      0.09826      0.17136      0.57      0.566
## Year2002     -0.08405      0.15827     -0.53      0.596
## Year2003     -0.02638      0.17571     -0.15      0.881
## Year2004      0.13714      0.15560      0.88      0.378
## Year2005      0.14972      0.13311      1.12      0.261
## Year2006     -0.07922      0.12981     -0.61      0.542
## Year2007      0.13273      0.13553      0.98      0.328
## Year2008     -0.12247      0.14447     -0.85      0.397
## Year2009      0.03243      0.12830      0.25      0.800
## Year2010      0.04420      0.13205      0.33      0.738
## Year2011     -0.24223      0.12452     -1.95      0.052 .
## Year2012     -0.10869      0.12234     -0.89      0.375
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.641
## Multiple R-squared:  0.259, Adjusted R-squared:  0.243
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 973 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0189 0.8720 0.9450 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.61e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.092 1          1.045
## LastAuthorFemale 1.118 1          1.057
## Year      1.094 16          1.003

```

## Residuals from first and last author



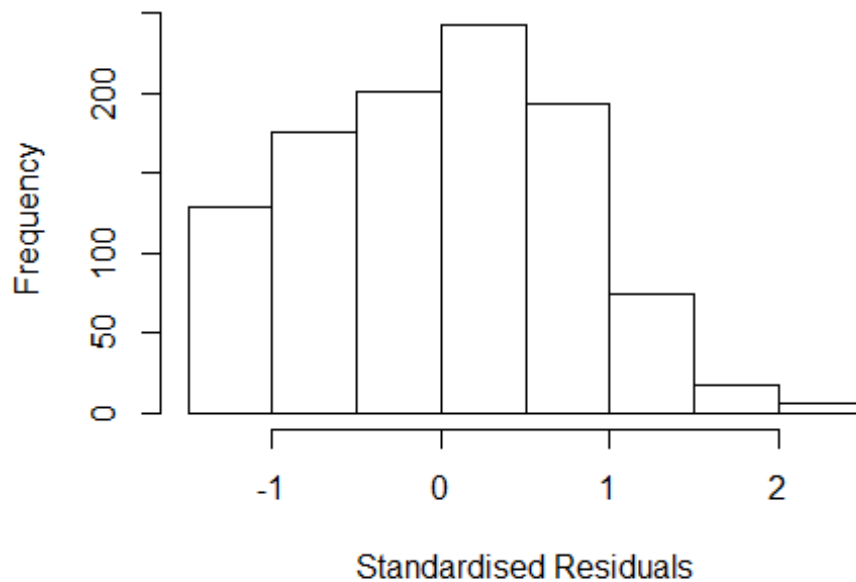
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3292 -0.6255 0.0344 0.5498 2.4678
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0228 0.1208 8.47 <2e-16 ***
## FirstAuthorFemale1 0.0427 0.0685 0.62 0.534
## LastAuthorFemale1 -0.0488 0.0961 -0.51 0.612
## Year1997 0.1155 0.1697 0.68 0.496
## Year1998 0.0484 0.1685 0.29 0.774
## Year1999 0.1369 0.2607 0.52 0.600
## Year2000 0.0148 0.1608 0.09 0.927
## Year2001 0.0226 0.1890 0.12 0.905
## Year2002 -0.0621 0.1789 -0.35 0.729
## Year2003 0.1605 0.1841 0.87 0.384
## Year2004 0.1752 0.1643 1.07 0.286
## Year2005 0.2894 0.1564 1.85 0.065 .
```

```

## Year2006          0.0182      0.1575      0.12      0.908
## Year2007          0.2443      0.1627      1.50      0.133
## Year2008          0.0121      0.1538      0.08      0.937
## Year2009          0.2327      0.1393      1.67      0.095 .
## Year2010          0.3065      0.1509      2.03      0.043 *
## Year2011         -0.0962      0.1418     -0.68      0.497
## Year2012          0.0419      0.1391      0.30      0.763
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.824
## Multiple R-squared:  0.0246, Adjusted R-squared:  0.00747
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 958 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.350  0.875   0.946   0.920   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1          1.016
## Year              1.032 16          1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3411 -0.6177 0.0292 0.5522 2.4683
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0191 0.1208 8.44 <2e-16 ***
## FirstAuthorFemale1 0.0315 0.0672 0.47 0.639
## Year1997 0.1186 0.1699 0.70 0.485
## Year1998 0.0511 0.1684 0.30 0.762
## Year1999 0.1405 0.2608 0.54 0.590
## Year2000 0.0159 0.1606 0.10 0.921
## Year2001 0.0186 0.1895 0.10 0.922
## Year2002 -0.0591 0.1789 -0.33 0.741
## Year2003 0.1636 0.1839 0.89 0.374
## Year2004 0.1756 0.1645 1.07 0.286
## Year2005 0.2905 0.1565 1.86 0.064 .
## Year2006 0.0193 0.1577 0.12 0.903
```

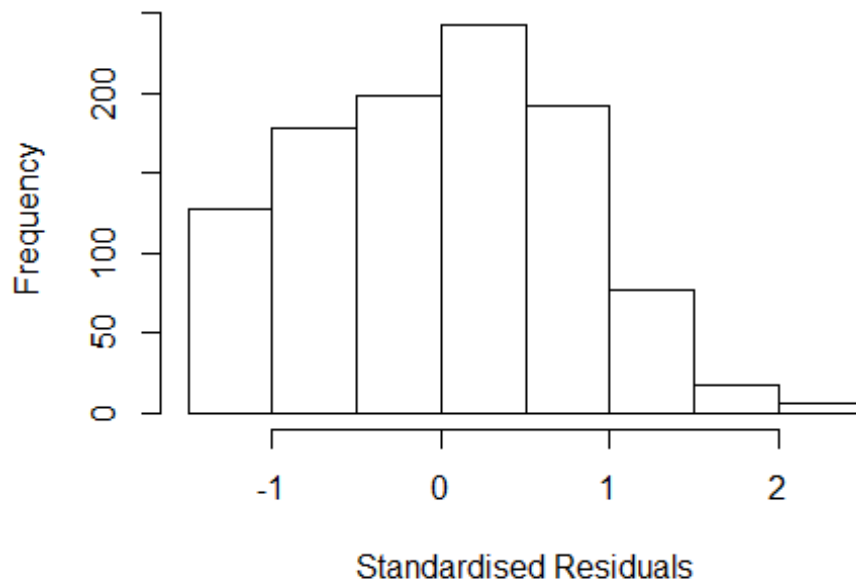
```

## Year2007          0.2458      0.1630      1.51      0.132
## Year2008          0.0119      0.1542      0.08      0.938
## Year2009          0.2347      0.1395      1.68      0.093 .
## Year2010          0.3069      0.1511      2.03      0.043 *
## Year2011         -0.0938      0.1420     -0.66      0.509
## Year2012          0.0447      0.1391      0.32      0.748
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.824
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.00819
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 956 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.349  0.874  0.946  0.920  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.057 1      1.028
## Year            1.057 16      1.002

```



## Residuals from last author



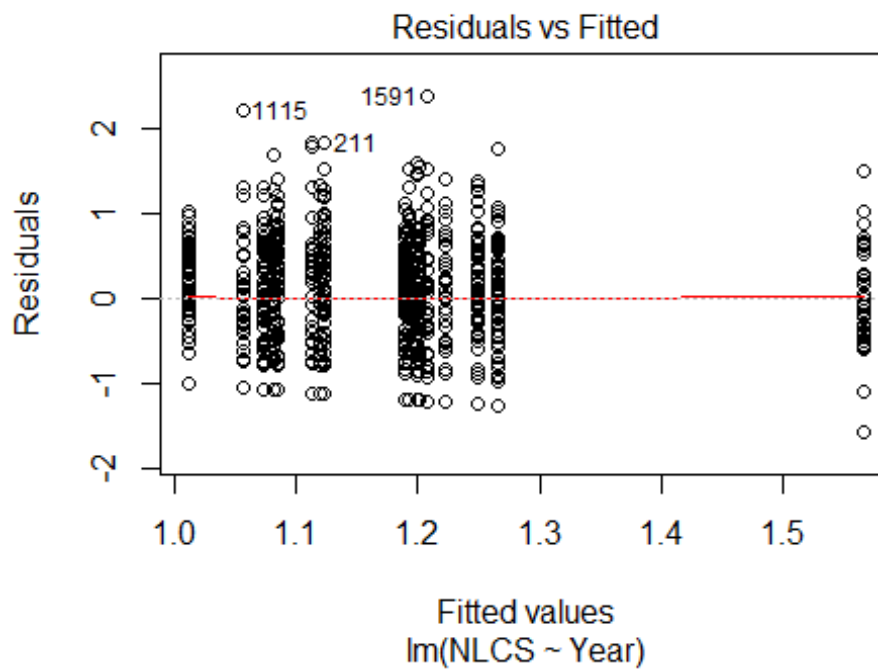
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3327 -0.6247 0.0484 0.5516 2.4626
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.02761 0.12011 8.56 <2e-16 ***
## LastAuthorFemale1 -0.03217 0.09302 -0.35 0.730
## Year1997 0.11574 0.16937 0.68 0.495
## Year1998 0.04738 0.16780 0.28 0.778
## Year1999 0.13197 0.26045 0.51 0.612
## Year2000 0.01175 0.16046 0.07 0.942
## Year2001 0.02336 0.18866 0.12 0.901
## Year2002 -0.05991 0.17836 -0.34 0.737
## Year2003 0.15752 0.18372 0.86 0.391
## Year2004 0.17799 0.16402 1.09 0.278
## Year2005 0.28682 0.15583 1.84 0.066 .
## Year2006 0.01682 0.15726 0.11 0.915
```

```

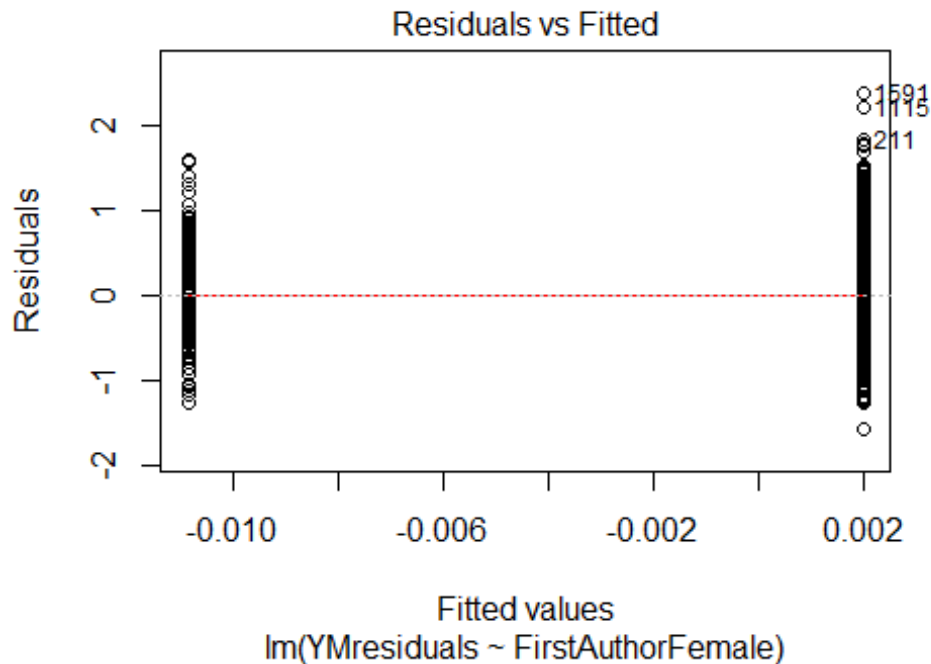
## Year2007      0.24536      0.16246      1.51      0.131
## Year2008      0.00966      0.15346      0.06      0.950
## Year2009      0.23371      0.13867      1.69      0.092 .
## Year2010      0.30513      0.15054      2.03      0.043 *
## Year2011     -0.09686      0.14142     -0.68      0.494
## Year2012      0.04345      0.13842      0.31      0.754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.823
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.00814
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 957 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.350  0.874  0.947  0.919  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1041"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2103"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 202 143 157 179 182 164 143 110 121 115 151 159 134 183 175
## 2011 2012
## 175 201
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62 41 44 40 38 33 51 43 43 39 62 68 58 91 61
## 2011 2012

```

```
## 70 84
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 35 38 37 34 28 45 37 38 35 53 56 48 75 50
## 2011 2012
## 61 74
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```

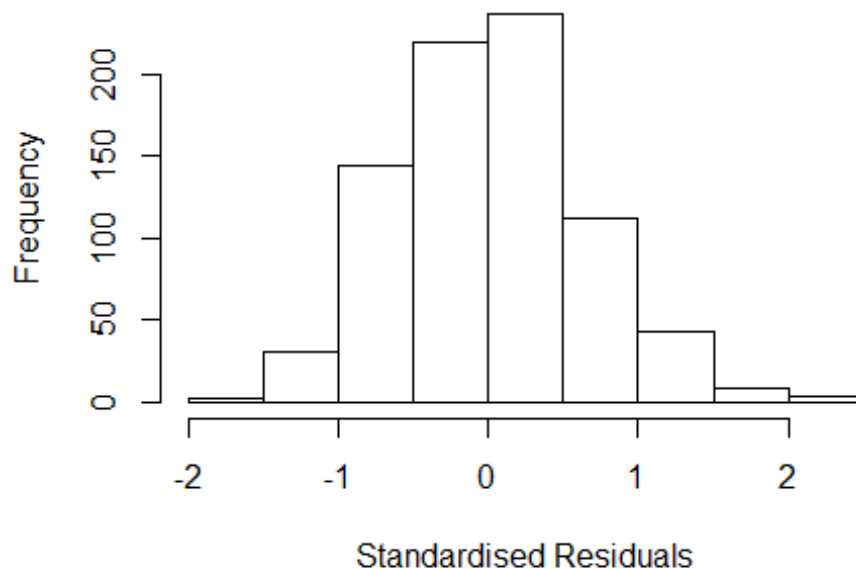


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.8328819311426"
## [1] "Male first author team size 2018 geometric mean: 2.8904127982457"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.09310167137661"
## [1] "Male last author team size 2018 geometric mean: 3.06857133471232"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 820, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.165 1 1.079
## LastAuthorFemale 1.292 1 1.137
## UniqueAuthors 1.380 4 1.041
## Year 1.683 16 1.016
```

## Residuals from first and last author and team size



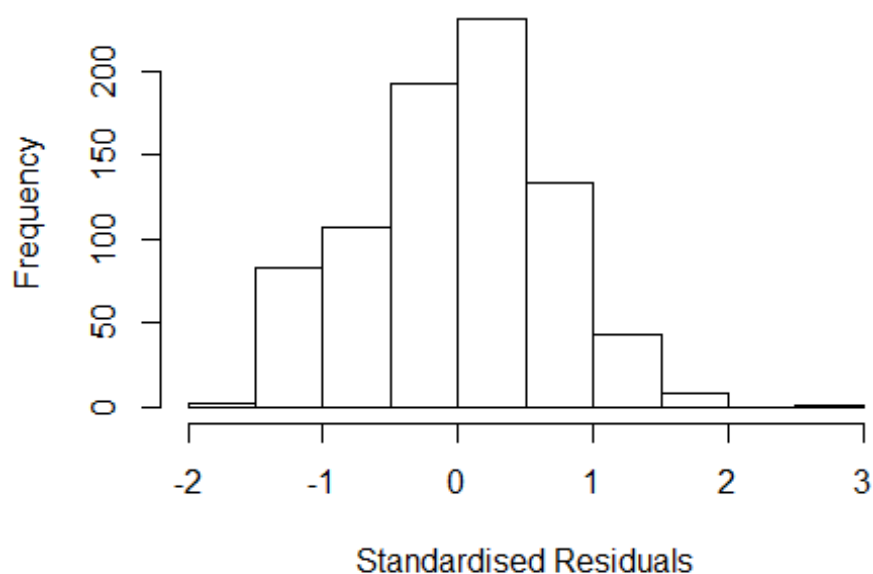
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.54145 -0.40263 0.00329 0.40836 2.34538
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6767 0.1025 6.60 7.6e-11 ***
## FirstAuthorFemale1 -0.1258 0.0620 -2.03 0.043 *
## LastAuthorFemale1 0.0658 0.0846 0.78 0.437
## UniqueAuthors2 0.5447 0.0720 7.57 1.1e-13 ***
## UniqueAuthors3 0.7129 0.0703 10.14 < 2e-16 ***
## UniqueAuthors4 0.6684 0.0827 8.08 2.5e-15 ***
## UniqueAuthors5 0.7340 0.0714 10.27 < 2e-16 ***
## Year1997 0.0582 0.1553 0.37 0.708
## Year1998 -0.0427 0.1399 -0.30 0.761
## Year1999 0.1518 0.1341 1.13 0.258
```

```

## Year2000          -0.0515      0.1722    -0.30      0.765
## Year2001          -0.0375      0.1474    -0.25      0.799
## Year2002          -0.0818      0.1681    -0.49      0.627
## Year2003          -0.0303      0.1511    -0.20      0.841
## Year2004           0.0192      0.1663      0.12      0.908
## Year2005           0.3192      0.1418      2.25      0.025 *
## Year2006           0.0385      0.1269      0.30      0.762
## Year2007           0.1352      0.1341      1.01      0.314
## Year2008          -0.1121      0.1361    -0.82      0.410
## Year2009           0.0352      0.1191      0.30      0.768
## Year2010          -0.0592      0.1224    -0.48      0.629
## Year2011          -0.1597      0.1135    -1.41      0.160
## Year2012          -0.1190      0.1142    -1.04      0.297
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.225, Adjusted R-squared:  0.203
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 728 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0713 0.8690 0.9450 0.8960 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.146 1      1.071
## LastAuthorFemale  1.182 1      1.087
## Year              1.205 16      1.006

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1591 2942628880 3.586 2004      2102      3      2.5
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5881 -0.4786  0.0305  0.4740  2.5001
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.054074   0.108132   9.75  < 2e-16 ***
## FirstAuthorFemale1 -0.014335   0.065342  -0.22  0.82640
## LastAuthorFemale1 -0.010713   0.089992  -0.12  0.90527
## Year1997         0.133427   0.162903   0.82  0.41300
## Year1998         0.042250   0.157282   0.27  0.78829
## Year1999         0.187155   0.155852   1.20  0.23017
## Year2000        -0.099820   0.169712  -0.59  0.55659
## Year2001        -0.126507   0.160329  -0.79  0.43033
## Year2002        -0.014246   0.176831  -0.08  0.93581
## Year2003         0.037628   0.159059   0.24  0.81306
## Year2004         0.031796   0.176752   0.18  0.85729
## Year2005         0.533988   0.155320   3.44  0.00062 ***
```

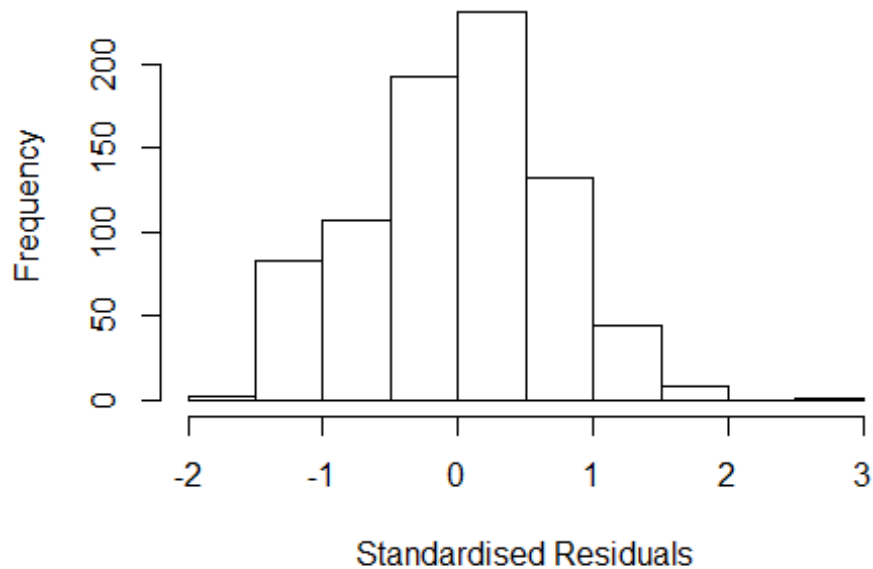
```

## Year2006          0.191065    0.144040    1.33  0.18507
## Year2007          0.220448    0.149440    1.48  0.14057
## Year2008         -0.051212    0.153943   -0.33  0.73947
## Year2009          0.161545    0.129857    1.24  0.21387
## Year2010          0.133979    0.138786    0.97  0.33466
## Year2011         -0.061104    0.132951   -0.46  0.64593
## Year2012         -0.000622    0.135263    0.00  0.99633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.689
## Multiple R-squared:  0.0404, Adjusted R-squared:  0.0183
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.160  0.856  0.951  0.911  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0        1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.081 1         1.040
## Year              1.081 16         1.002

```



## Residuals from first author



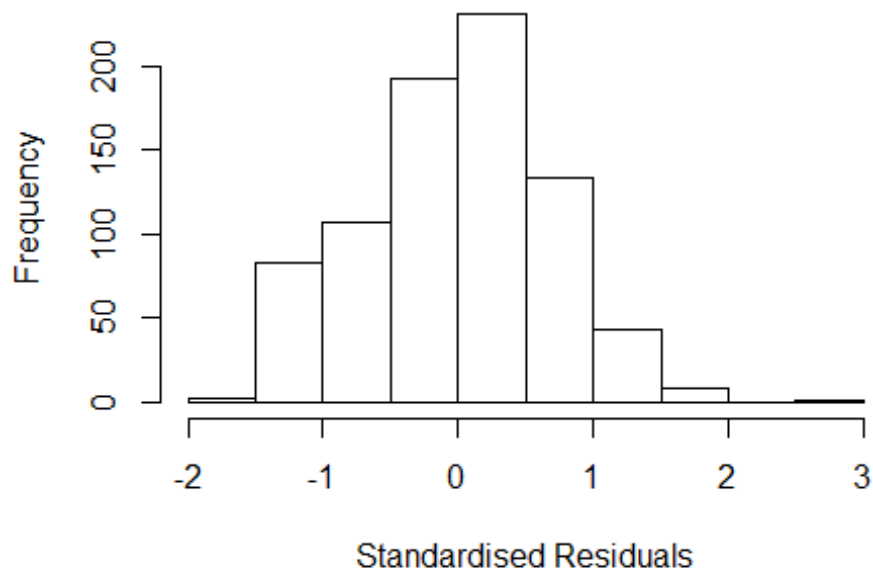
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1591 2942628880 3.586 2004      2102      3      2.5
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5871 -0.4802  0.0306  0.4746  2.5012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.053213   0.108592   9.70  < 2e-16 ***
## FirstAuthorFemale1 -0.017250   0.063804  -0.27  0.78695
## Year1997         0.135144   0.163151   0.83  0.40773
## Year1998         0.043318   0.157700   0.27  0.78363
## Year1999         0.188065   0.155983   1.21  0.22831
## Year2000        -0.099639   0.169967  -0.59  0.55789
## Year2001        -0.128205   0.159145  -0.81  0.42072
## Year2002        -0.013944   0.177721  -0.08  0.93748
## Year2003         0.038009   0.159384   0.24  0.81158
## Year2004         0.031550   0.177238   0.18  0.85876
## Year2005         0.533913   0.155241   3.44  0.00061 ***
## Year2006         0.191367   0.144339   1.33  0.18529
```

```

## Year2007          0.220764    0.149769    1.47  0.14087
## Year2008          -0.051321    0.154068   -0.33  0.73914
## Year2009          0.162436    0.130186    1.25  0.21251
## Year2010          0.133901    0.138944    0.96  0.33549
## Year2011          -0.060215    0.133418   -0.45  0.65188
## Year2012          0.000314    0.135579    0.00  0.99815
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.0197
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.153  0.852  0.950  0.909  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.108 1          1.053
## Year            1.108 16          1.003

```

## Residuals from last author



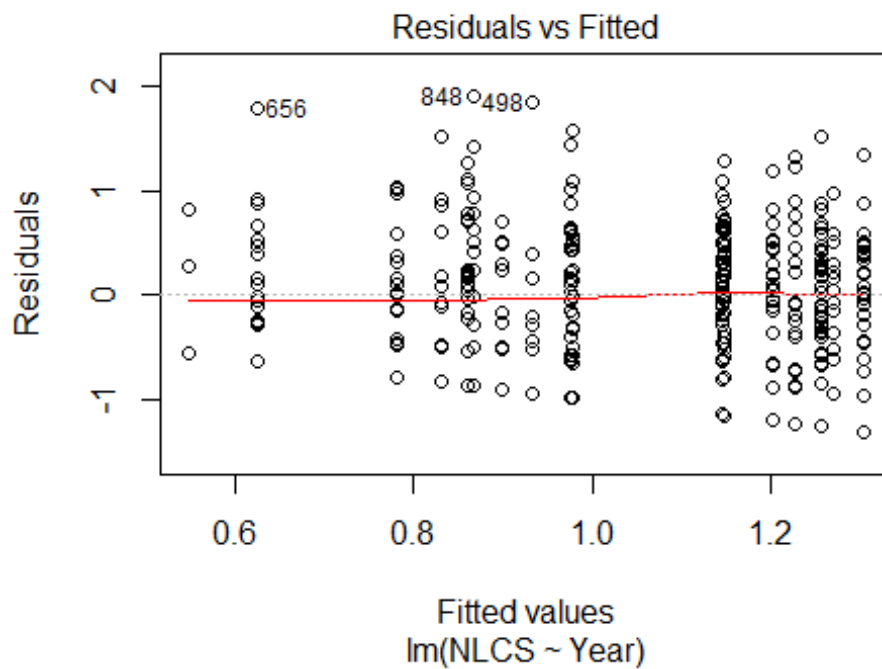
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1591 2942628880 3.586 2004      2102      3      2.5
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5861 -0.4833  0.0318  0.4749  2.5019
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05258    0.10796   9.75  < 2e-16 ***
## LastAuthorFemale1 -0.01650    0.08773  -0.19  0.85086
## Year1997         0.13153    0.16254   0.81  0.41866
## Year1998         0.04179    0.15744   0.27  0.79076
## Year1999         0.18695    0.15579   1.20  0.23051
## Year2000        -0.09786    0.16919  -0.58  0.56316
## Year2001        -0.12658    0.16056  -0.79  0.43070
## Year2002        -0.01625    0.17615  -0.09  0.92653
## Year2003         0.03839    0.15924   0.24  0.80956
## Year2004         0.03155    0.17635   0.18  0.85807
## Year2005         0.53348    0.15550   3.43  0.00063 ***
## Year2006         0.19023    0.14403   1.32  0.18696
```

```

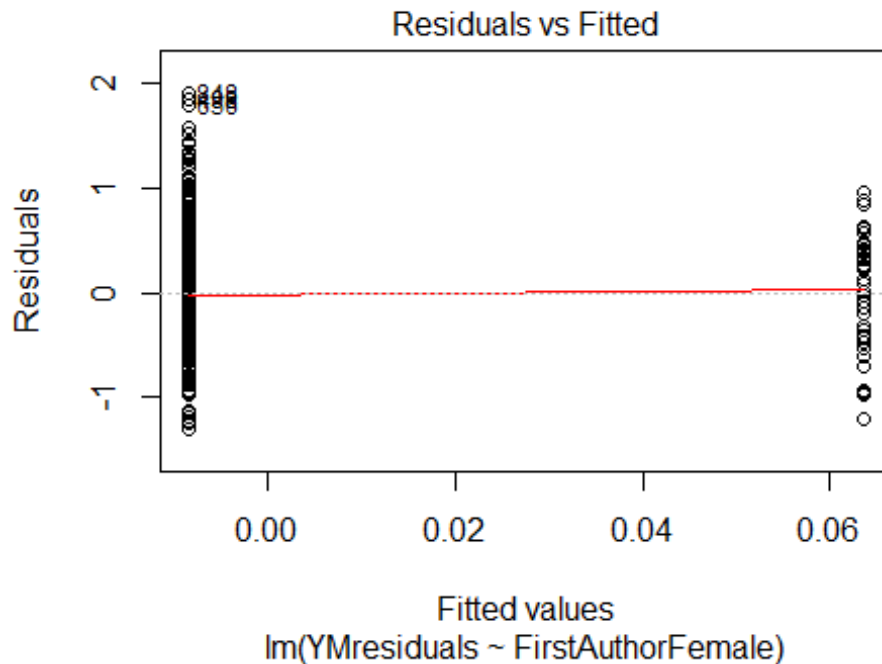
## Year2007          0.21991      0.14938      1.47  0.14139
## Year2008          -0.05103     0.15397     -0.33  0.74041
## Year2009          0.16051     0.12978      1.24  0.21655
## Year2010          0.13490     0.13882      0.97  0.33148
## Year2011          -0.06115     0.13306     -0.46  0.64595
## Year2012          -0.00236     0.13486     -0.02  0.98602
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.693
## Multiple R-squared:  0.0402, Adjusted R-squared:  0.0194
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.165  0.859  0.950  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 800"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2104"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 116 136 97 103 121 85 59 52 80 48 97 82 101 125 103
## 2011 2012
## 130 116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 23 14 13 8 26 4 19 18 19 15 34 21 25 37 32
## 2011 2012

```

```
## 44 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 12 13 7 24 4 17 16 15 10 25 14 19 30 28
## 2011 2012
## 36 37
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 11, df = 16, p-value = 0.8
```

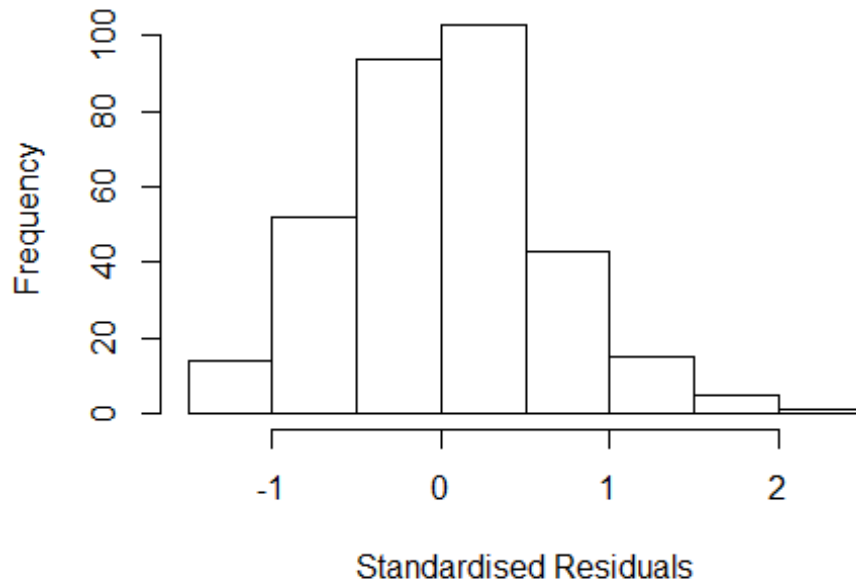


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.84420392254925"
## [1] "Male first author team size 2018 geometric mean: 3.21750551922208"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 810, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.20962000380523"
## [1] "Male last author team size 2018 geometric mean: 3.36500523079308"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 410, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.419  1          1.191
## LastAuthorFemale  1.274  1          1.129
## UniqueAuthors    3.241  4          1.158
## Year              3.165 16          1.037
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3987 -0.4501 0.0224 0.3879 2.0936
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8281 0.1414 5.85 1.2e-08 ***
## FirstAuthorFemale1 -0.0464 0.1156 -0.40 0.6882
## LastAuthorFemale1 0.0761 0.1799 0.42 0.6724
## UniqueAuthors2 0.3052 0.1071 2.85 0.0047 **
## UniqueAuthors3 0.3481 0.1109 3.14 0.0019 **
## UniqueAuthors4 0.4158 0.1289 3.23 0.0014 **
## UniqueAuthors5 0.6156 0.1393 4.42 1.4e-05 ***
## Year1997 -0.2768 0.2391 -1.16 0.2479
## Year1998 -0.0541 0.2098 -0.26 0.7966
## Year1999 -0.1547 0.2819 -0.55 0.5835
```

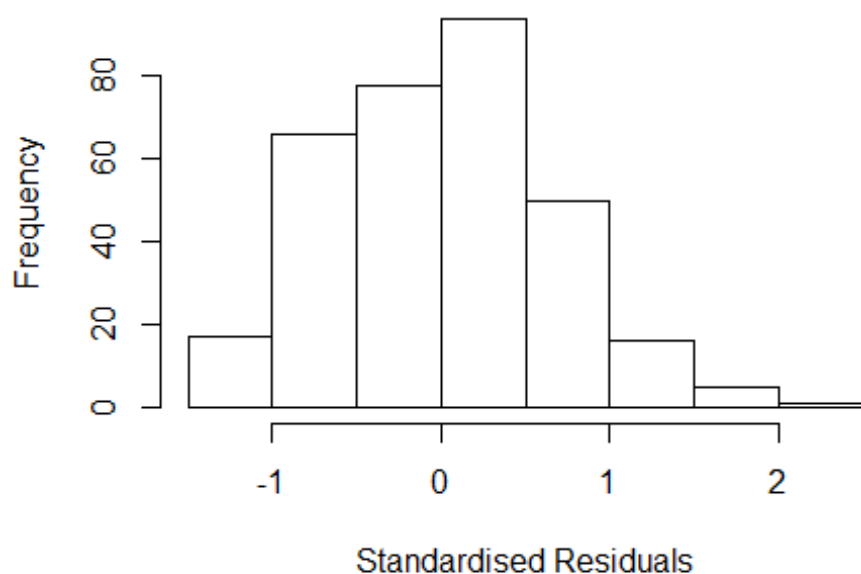
```

## Year2000          -0.3684      0.1876    -1.96    0.0505 .
## Year2001          -0.4399      0.3499    -1.26    0.2096
## Year2002          -0.4603      0.2806    -1.64    0.1020
## Year2003          -0.2018      0.2104    -0.96    0.3381
## Year2004           0.2205      0.2929     0.75    0.4522
## Year2005           0.1629      0.2071     0.79    0.4322
## Year2006          -0.1572      0.2112    -0.74    0.4572
## Year2007          -0.4732      0.2018    -2.35    0.0197 *
## Year2008          -0.0150      0.2000    -0.08    0.9401
## Year2009          -0.0449      0.1981    -0.23    0.8207
## Year2010           0.1341      0.1911     0.70    0.4832
## Year2011           0.0346      0.1805     0.19    0.8482
## Year2012           0.0749      0.1836     0.41    0.6835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.198, Adjusted R-squared:  0.14
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 302 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.223  0.873  0.951  0.908  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.273 1      1.128
## LastAuthorFemale  1.189 1      1.090
## Year              1.503 16      1.013

```



## Residuals from first and last author

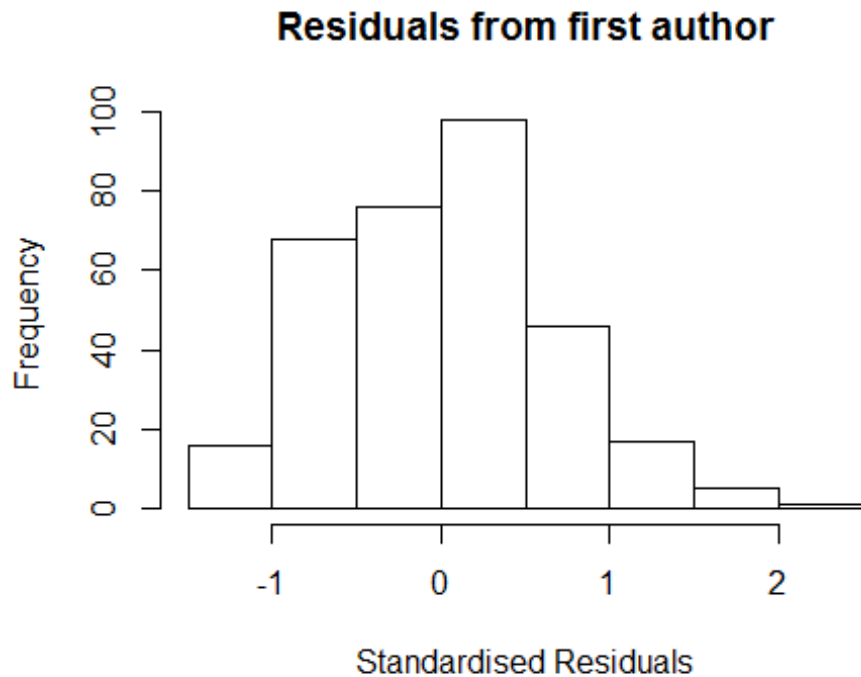


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3189 -0.5066 0.0121 0.4585 2.1005
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90957 0.16113 5.64 3.8e-08 ***
## FirstAuthorFemale1 0.07532 0.10874 0.69 0.489
## LastAuthorFemale1 0.07521 0.17466 0.43 0.667
## Year1997 -0.25420 0.25263 -1.01 0.315
## Year1998 -0.00199 0.24086 -0.01 0.993
## Year1999 -0.08494 0.30233 -0.28 0.779
## Year2000 -0.31145 0.20131 -1.55 0.123
## Year2001 -0.37733 0.36879 -1.02 0.307
## Year2002 -0.24607 0.28661 -0.86 0.391
## Year2003 -0.05079 0.23943 -0.21 0.832
## Year2004 0.32832 0.28308 1.16 0.247
## Year2005 0.32321 0.20831 1.55 0.122
```

```

## Year2006      -0.02106    0.21621   -0.10    0.922
## Year2007      -0.30829    0.21193   -1.45    0.147
## Year2008       0.13201    0.21368    0.62    0.537
## Year2009       0.21171    0.19578    1.08    0.280
## Year2010       0.40932    0.20256    2.02    0.044 *
## Year2011       0.30286    0.19932    1.52    0.130
## Year2012       0.32304    0.18454    1.75    0.081 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.135, Adjusted R-squared:  0.084
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 304 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.248  0.894  0.944  0.906  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.266 1      1.125
## Year              1.266 16      1.007

```



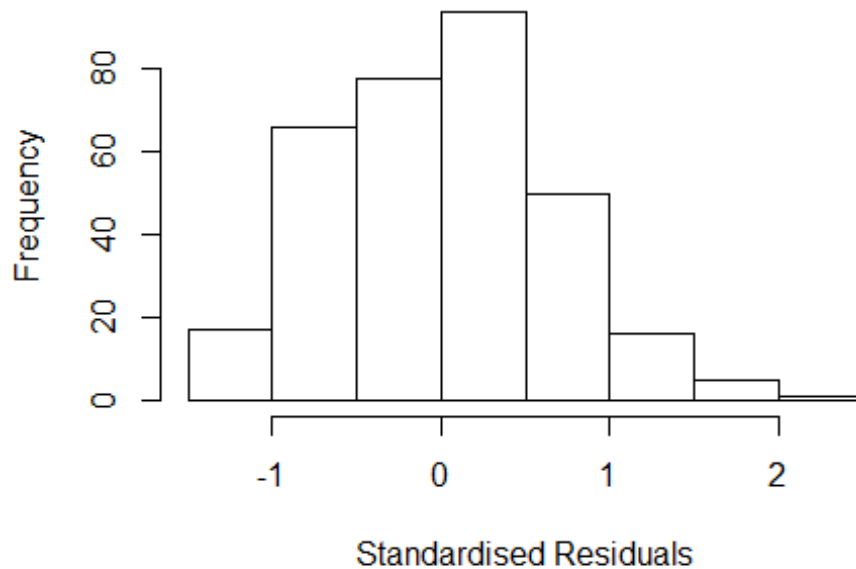
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3246 -0.5098  0.0105  0.4525  2.0868
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91631    0.15899   5.76   2e-08 ***
## FirstAuthorFemale1 0.08015    0.10916   0.73   0.463
## Year1997      -0.26084    0.25132  -1.04   0.300
## Year1998      -0.00655    0.23664  -0.03   0.978
## Year1999      -0.09122    0.30148  -0.30   0.762
## Year2000      -0.31820    0.19970  -1.59   0.112
## Year2001      -0.38402    0.36782  -1.04   0.297
## Year2002      -0.23914    0.28475  -0.84   0.402
## Year2003      -0.05811    0.23796  -0.24   0.807
## Year2004       0.32478    0.27893   1.16   0.245
## Year2005       0.31952    0.20854   1.53   0.126
## Year2006      -0.01840    0.21617  -0.09   0.932
```

```

## Year2007      -0.30968    0.21228   -1.46    0.146
## Year2008      0.13399    0.21138    0.63    0.527
## Year2009      0.21106    0.19387    1.09    0.277
## Year2010      0.40827    0.20033    2.04    0.042 *
## Year2011      0.29814    0.19795    1.51    0.133
## Year2012      0.31914    0.18377    1.74    0.083 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.133, Adjusted R-squared:  0.0855
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 305 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.895  0.944  0.906  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.181 1      1.087
## Year      1.181 16      1.005

```

## Residuals from last author



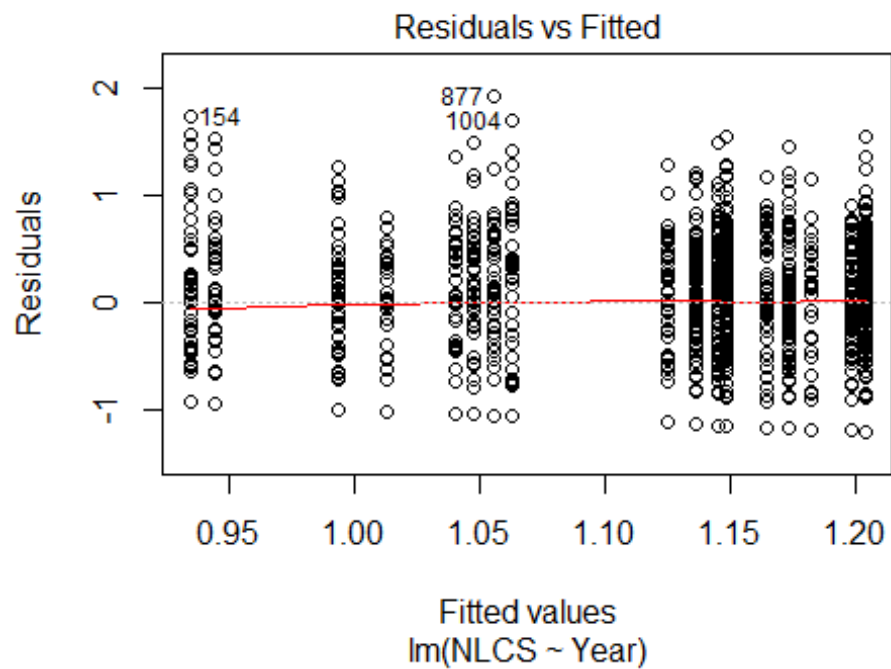
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33414 -0.50824 0.00689 0.46367 2.09452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9184 0.1582 5.80 1.6e-08 ***
## LastAuthorFemale1 0.0838 0.1740 0.48 0.630
## Year1997 -0.2631 0.2510 -1.05 0.295
## Year1998 -0.0054 0.2344 -0.02 0.982
## Year1999 -0.0937 0.3013 -0.31 0.756
## Year2000 -0.3183 0.2001 -1.59 0.113
## Year2001 -0.3862 0.3677 -1.05 0.294
## Year2002 -0.2490 0.2864 -0.87 0.385
## Year2003 -0.0478 0.2412 -0.20 0.843
## Year2004 0.3264 0.2815 1.16 0.247
## Year2005 0.3223 0.2065 1.56 0.120
## Year2006 -0.0177 0.2153 -0.08 0.935
```

```

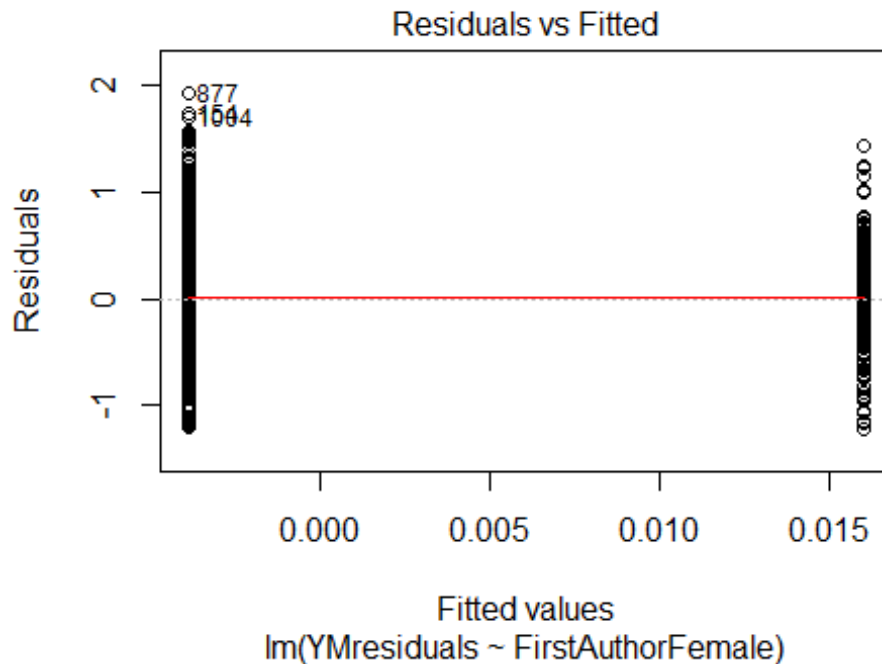
## Year2007          -0.3178      0.2097    -1.52      0.131
## Year2008           0.1303      0.2122      0.61      0.540
## Year2009           0.2127      0.1958      1.09      0.278
## Year2010           0.4157      0.2024      2.05      0.041 *
## Year2011           0.3037      0.1995      1.52      0.129
## Year2012           0.3315      0.1842      1.80      0.073 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.133, Adjusted R-squared:  0.0858
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 308 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.251  0.898  0.945  0.907  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 327"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2105"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 184 108 137 118 145 104 97 81 119 103 133 197 185 230 255
## 2011 2012
## 294 311
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 33 44 42 47 43 45 30 50 51 58 83 91 99 121
## 2011 2012

```

```
## 157 175
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 29 39 41 42 37 42 28 47 48 52 78 77 82 110
## 2011 2012
## 131 159
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



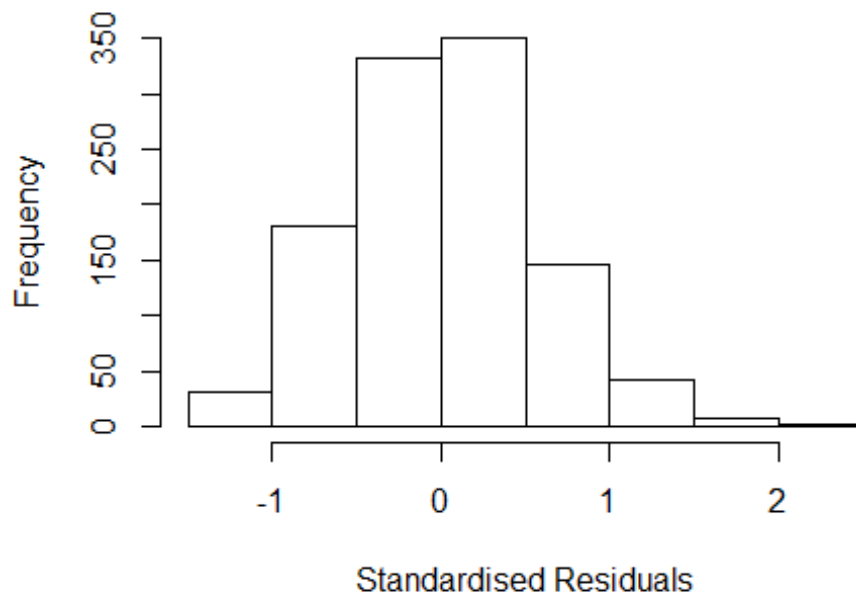
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.1, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 2.89864328591439"
## [1] "Male first author team size 2018 geometric mean: 3.20070744431243"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.61011183797719"
## [1] "Male last author team size 2018 geometric mean: 3.24903299269954"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8500, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.474 1      1.214
## LastAuthorFemale  1.445 1      1.202
## UniqueAuthors    1.540 4      1.055
## Year             1.669 16      1.016
```



## Residuals from first and last author and team size



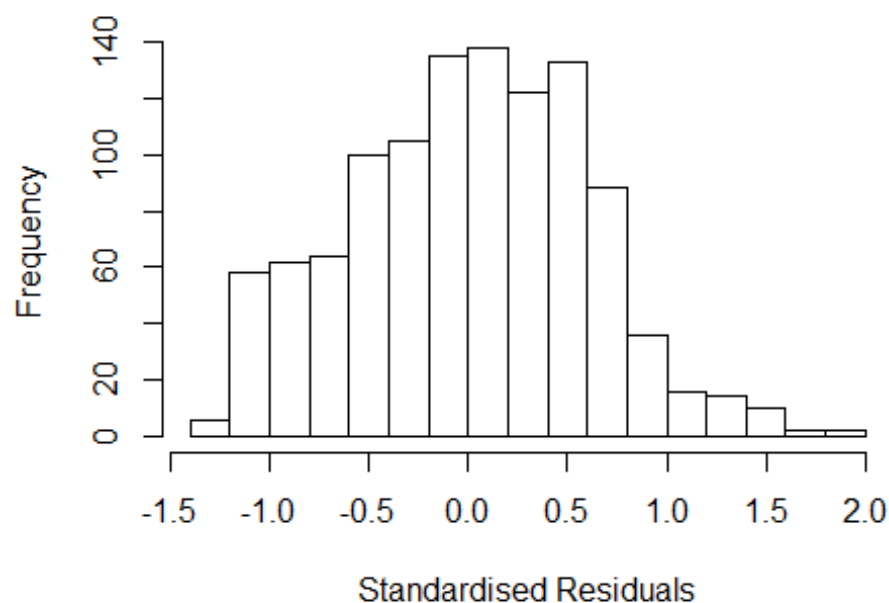
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35798 -0.38285 0.00187 0.37035 2.05568
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6558 0.0989 6.63 5.2e-11 ***
## FirstAuthorFemale1 -0.0644 0.0497 -1.30 0.1951
## LastAuthorFemale1 0.1421 0.0582 2.44 0.0148 *
## UniqueAuthors2 0.3836 0.0500 7.67 3.7e-14 ***
## UniqueAuthors3 0.4964 0.0544 9.12 < 2e-16 ***
## UniqueAuthors4 0.4507 0.0627 7.18 1.3e-12 ***
## UniqueAuthors5 0.5591 0.0556 10.06 < 2e-16 ***
## Year1997 0.2206 0.1335 1.65 0.0988 .
## Year1998 0.0399 0.1475 0.27 0.7867
## Year1999 0.1739 0.1420 1.22 0.2210
```

```

## Year2000          0.1366      0.1370      1.00      0.3189
## Year2001          0.1238      0.1426      0.87      0.3853
## Year2002          0.0485      0.1872      0.26      0.7955
## Year2003          0.0848      0.1421      0.60      0.5507
## Year2004          0.1083      0.1316      0.82      0.4105
## Year2005          0.1325      0.1354      0.98      0.3279
## Year2006          0.2043      0.1309      1.56      0.1190
## Year2007          0.3077      0.1152      2.67      0.0077 **
## Year2008          0.2377      0.1144      2.08      0.0380 *
## Year2009          0.1798      0.1157      1.55      0.1206
## Year2010          0.2058      0.1123      1.83      0.0672 .
## Year2011          0.0966      0.1131      0.85      0.3933
## Year2012          0.1855      0.1088      1.71      0.0884 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.162, Adjusted R-squared:  0.145
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 78 weights are ~= 1. The remaining 1013 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.149  0.867  0.953  0.907  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.392 1      1.180
## LastAuthorFemale  1.372 1      1.172
## Year              1.100 16      1.003

```

## Residuals from first and last author



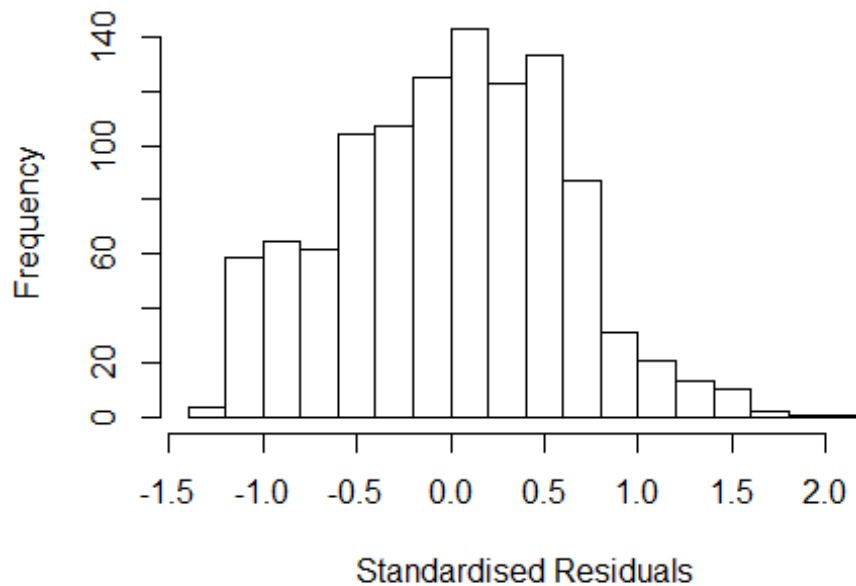
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2830 -0.4383 0.0211 0.4386 1.9487
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.79627 0.10484 7.59 6.7e-14 ***
## FirstAuthorFemale1 0.00759 0.05031 0.15 0.8801
## LastAuthorFemale1 0.10012 0.05931 1.69 0.0917 .
## Year1997 0.32532 0.14991 2.17 0.0302 *
## Year1998 0.08507 0.16314 0.52 0.6021
## Year1999 0.25063 0.15196 1.65 0.0994 .
## Year2000 0.26363 0.14176 1.86 0.0632 .
## Year2001 0.13489 0.15277 0.88 0.3775
## Year2002 0.18299 0.19939 0.92 0.3589
## Year2003 0.20735 0.14337 1.45 0.1484
## Year2004 0.17619 0.13615 1.29 0.1959
## Year2005 0.21509 0.14496 1.48 0.1382
```

```

## Year2006          0.34684    0.14260    2.43    0.0152 *
## Year2007          0.42173    0.11878    3.55    0.0004 ***
## Year2008          0.31380    0.12225    2.57    0.0104 *
## Year2009          0.33315    0.12437    2.68    0.0075 **
## Year2010          0.34081    0.12013    2.84    0.0046 **
## Year2011          0.32761    0.11733    2.79    0.0053 **
## Year2012          0.37900    0.11609    3.26    0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0152
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 980 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.285  0.872  0.945  0.908  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## Year              1.044 16      1.001

```

## Residuals from first author



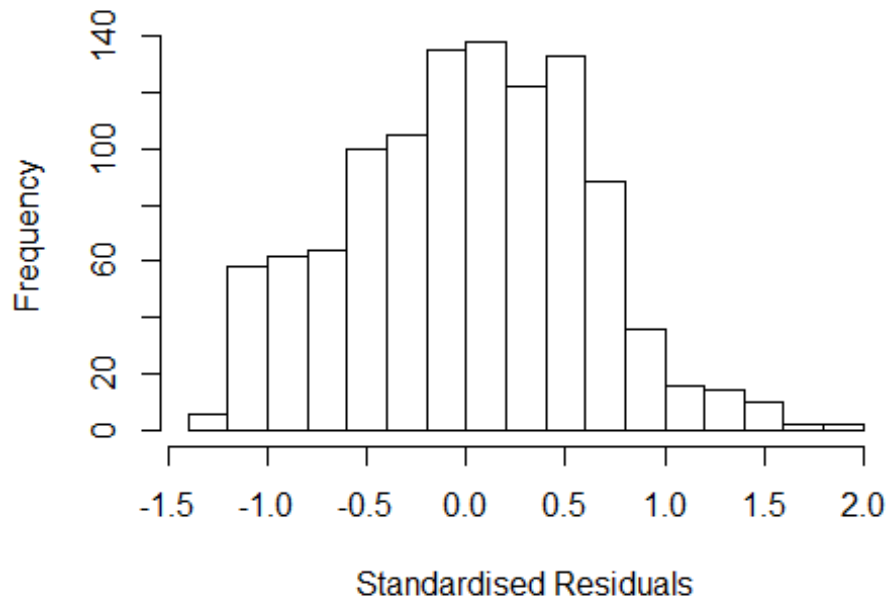
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2351 -0.4471  0.0219  0.4436  2.0400
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8068    0.1052   7.67 3.8e-14 ***
## FirstAuthorFemale1 0.0491    0.0432   1.13 0.25697
## Year1997        0.3184    0.1514   2.10 0.03573 *
## Year1998        0.0892    0.1649   0.54 0.58863
## Year1999        0.2425    0.1523   1.59 0.11170
## Year2000        0.2626    0.1429   1.84 0.06634 .
## Year2001        0.1332    0.1504   0.89 0.37593
## Year2002        0.1772    0.2005   0.88 0.37682
## Year2003        0.2000    0.1439   1.39 0.16495
## Year2004        0.1686    0.1366   1.23 0.21745
## Year2005        0.2082    0.1455   1.43 0.15271
## Year2006        0.3512    0.1438   2.44 0.01475 *
```

```

## Year2007          0.4154      0.1191      3.49  0.00051 ***
## Year2008          0.3131      0.1226      2.55  0.01078 *
## Year2009          0.3227      0.1238      2.61  0.00928 **
## Year2010          0.3373      0.1205      2.80  0.00522 **
## Year2011          0.3222      0.1175      2.74  0.00621 **
## Year2012          0.3792      0.1166      3.25  0.00118 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.0288, Adjusted R-squared:  0.0134
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 983 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.239  0.876  0.944  0.908  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1      1.016
## Year              1.032 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2799 -0.4380 0.0207 0.4379 1.9435
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7970 0.1048 7.61 6.1e-14 ***
## LastAuthorFemale1 0.1040 0.0515 2.02 0.0434 *
## Year1997 0.3257 0.1496 2.18 0.0298 *
## Year1998 0.0846 0.1629 0.52 0.6034
## Year1999 0.2507 0.1519 1.65 0.0990 .
## Year2000 0.2634 0.1418 1.86 0.0635 .
## Year2001 0.1355 0.1527 0.89 0.3752
## Year2002 0.1832 0.1991 0.92 0.3578
## Year2003 0.2073 0.1433 1.45 0.1484
## Year2004 0.1766 0.1363 1.30 0.1953
## Year2005 0.2156 0.1449 1.49 0.1370
## Year2006 0.3472 0.1423 2.44 0.0148 *
```

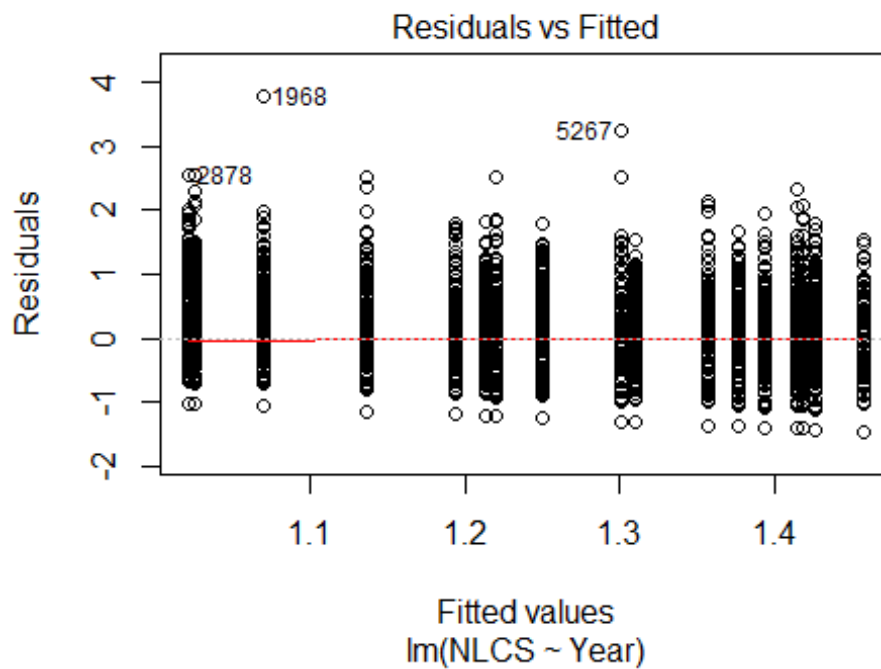
```

## Year2007          0.4217      0.1187      3.55      0.0004 ***
## Year2008          0.3140      0.1223      2.57      0.0104 *
## Year2009          0.3340      0.1238      2.70      0.0071 **
## Year2010          0.3413      0.1200      2.84      0.0045 **
## Year2011          0.3283      0.1170      2.81      0.0051 **
## Year2012          0.3789      0.1161      3.26      0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.0314, Adjusted R-squared:  0.0161
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 977 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.289  0.871  0.944  0.908  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1091"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2200"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1139 848 808 725 563 411 306 332 297 346 429 417 399 484 441
## 2011 2012
## 431 410
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 348 228 245 251 217 129 113 155 118 175 217 190 176 256 204
## 2011 2012

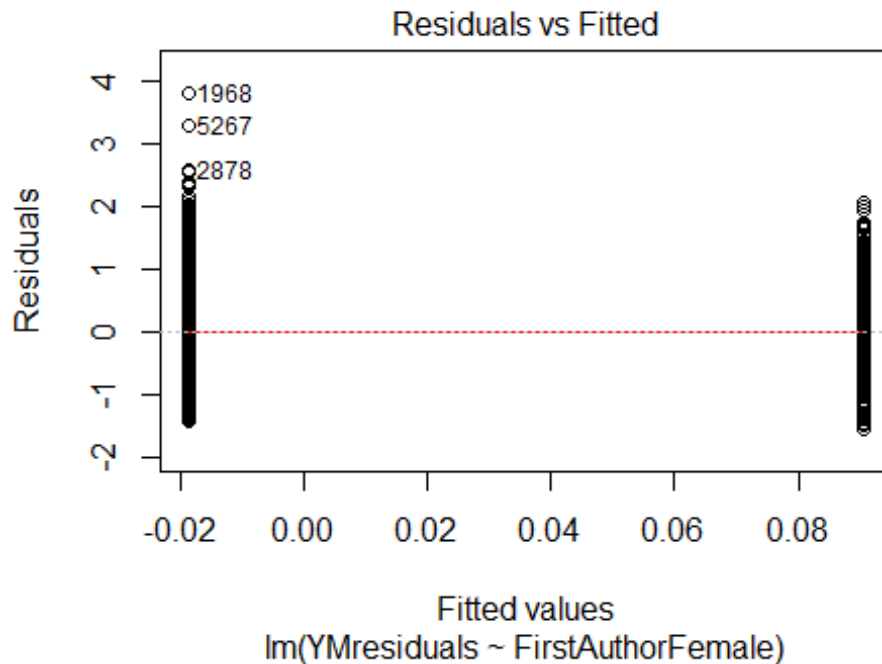
```



```
## 216 224
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 330 213 220 225 194 122 101 139 108 163 194 160 156 227 174
## 2011 2012
## 186 192
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 67, df = 16, p-value = 4e-08
```

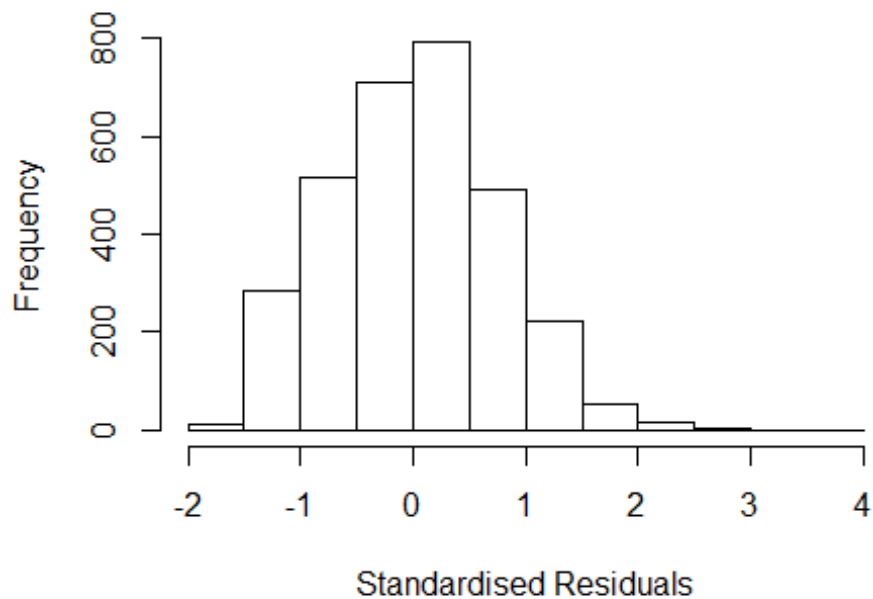


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.7, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 2.92972819556116"
## [1] "Male first author team size 2018 geometric mean: 2.40722321410961"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.39628272839137"
## [1] "Male last author team size 2018 geometric mean: 2.5277639037288"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.181 1          1.087
## LastAuthorFemale  1.155 1          1.075
## UniqueAuthors    1.215 4          1.025
## Year              1.225 16         1.006
```

## Residuals from first and last author and team size



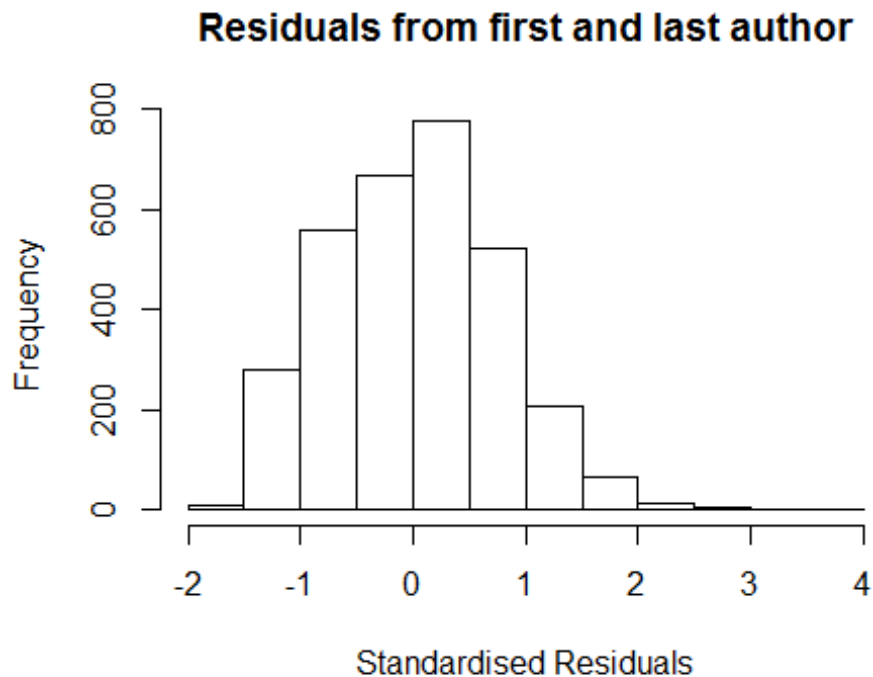
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1063  0007244732 3.574 1996    2200      3    2.791
## 1968  0031119324 4.853 1997    1210      2    3.708
## 4328  0346041912 3.749 2000    2200      1    2.709
## 5267  0036565284 4.562 2002    1210      2    3.247
## 6298  21844442335 3.755 2005    2200      1    2.554
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6375 -0.5398  0.0204  0.5093  3.7085
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7833    0.0500   15.68 < 2e-16 ***
## FirstAuthorFemale1 0.0670    0.0376    1.78  0.07461 .
## LastAuthorFemale1 0.0577    0.0441    1.31  0.19120
## UniqueAuthors2    0.2482    0.0360    6.90  6.1e-12 ***
## UniqueAuthors3    0.3191    0.0400    7.97  2.2e-15 ***
## UniqueAuthors4    0.4802    0.0512    9.38 < 2e-16 ***
```

```

## UniqueAuthors5      0.4483      0.0516      8.69 < 2e-16 ***
## Year1997             0.0422      0.0780      0.54 0.58904
## Year1998            -0.0430      0.0808     -0.53 0.59490
## Year1999             0.1682      0.0746      2.25 0.02430 *
## Year2000             0.2565      0.0716      3.59 0.00034 ***
## Year2001             0.3427      0.0924      3.71 0.00021 ***
## Year2002             0.2834      0.0914      3.10 0.00195 **
## Year2003             0.3795      0.0746      5.09 3.8e-07 ***
## Year2004             0.4315      0.0724      5.96 2.8e-09 ***
## Year2005             0.4176      0.0744      5.62 2.1e-08 ***
## Year2006             0.3741      0.0676      5.53 3.4e-08 ***
## Year2007             0.4045      0.0703      5.75 9.5e-09 ***
## Year2008             0.2171      0.0755      2.87 0.00408 **
## Year2009             0.2832      0.0658      4.31 1.7e-05 ***
## Year2010             0.1676      0.0712      2.35 0.01867 *
## Year2011             0.1486      0.0707      2.10 0.03570 *
## Year2012             0.3289      0.0669      4.92 9.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.76
## Multiple R-squared:  0.0938, Adjusted R-squared:  0.0873
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 486 is an outlier with |weight| = 0 ( < 3.2e-05);
## 262 weights are ~= 1. The remaining 2841 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0284 0.8680 0.9470 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.22e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.145 1          1.070

```

```
## LastAuthorFemale  1.136  1          1.066
## Year              1.036 16          1.001
```



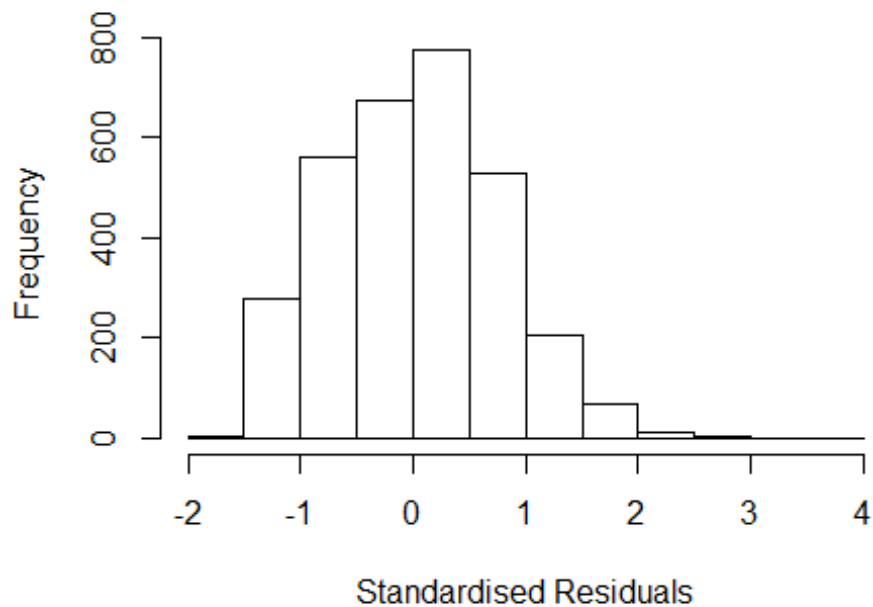
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1063 0007244732 3.574 1996    2200      3    2.596
## 1968 0031119324 4.853 1997    1210      2    3.862
## 2878 0031599055 3.579 1998    1405      2    2.684
## 4328 0346041912 3.749 2000    2200      1    2.555
## 5052 0036805915 3.818 2002    1210      2    2.541
## 5267 0036565284 4.562 2002    1210      2    3.285
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5686 -0.5478  0.0237  0.5268  3.8615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9778     0.0485   20.14 < 2e-16 ***
## FirstAuthorFemale1 0.1131     0.0377    3.00  0.00273 **
## LastAuthorFemale1 0.0393     0.0449    0.88  0.38067
## Year1997        0.0137     0.0803    0.17  0.86489
```

```

## Year1998          -0.0832      0.0835    -1.00    0.31913
## Year1999          0.1415      0.0768     1.84    0.06542 .
## Year2000          0.2159      0.0746     2.89    0.00383 **
## Year2001          0.3414      0.0951     3.59    0.00034 ***
## Year2002          0.2992      0.0909     3.29    0.00101 **
## Year2003          0.3840      0.0768     5.00    6.0e-07 ***
## Year2004          0.4383      0.0752     5.83    6.2e-09 ***
## Year2005          0.4152      0.0779     5.33    1.1e-07 ***
## Year2006          0.3767      0.0716     5.26    1.6e-07 ***
## Year2007          0.3868      0.0723     5.35    9.5e-08 ***
## Year2008          0.1712      0.0789     2.17    0.03018 *
## Year2009          0.3403      0.0676     5.03    5.1e-07 ***
## Year2010          0.2097      0.0738     2.84    0.00454 **
## Year2011          0.2060      0.0738     2.79    0.00529 **
## Year2012          0.4039      0.0683     5.91    3.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.778
## Multiple R-squared:  0.0485, Adjusted R-squared:  0.043
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 486 is an outlier with |weight| = 0 ( < 3.2e-05);
## 265 weights are ~= 1. The remaining 2838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0355 0.8610 0.9480 0.9120 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.22e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1           1.013
## Year              1.026 16           1.001

```

## Residuals from first author



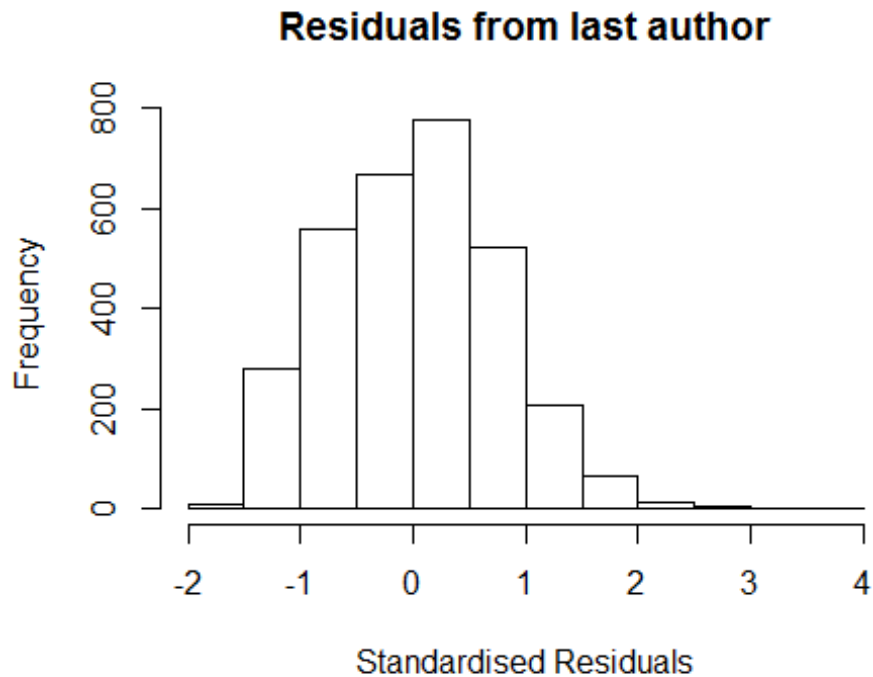
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1063 0007244732 3.574 1996    2200      3    2.596
## 1968 0031119324 4.853 1997    1210      2    3.862
## 2878 0031599055 3.579 1998    1405      2    2.684
## 4328 0346041912 3.749 2000    2200      1    2.555
## 5052 0036805915 3.818 2002    1210      2    2.541
## 5267 0036565284 4.562 2002    1210      2    3.285
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5449 -0.5500  0.0199  0.5281  3.8600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9803     0.0485   20.21 < 2e-16 ***
## FirstAuthorFemale1 0.1270     0.0358    3.55 0.00040 ***
## Year1997        0.0127     0.0802    0.16 0.87442
## Year1998       -0.0823     0.0834   -0.99 0.32366
## Year1999        0.1416     0.0768    1.84 0.06528 .
## Year2000        0.2158     0.0746    2.89 0.00384 **
## Year2001        0.3434     0.0952    3.61 0.00032 ***
```

```

## Year2002          0.3025      0.0908      3.33  0.00088 ***
## Year2003          0.3852      0.0769      5.01  5.7e-07 ***
## Year2004          0.4376      0.0753      5.81  6.8e-09 ***
## Year2005          0.4166      0.0780      5.34  9.9e-08 ***
## Year2006          0.3777      0.0716      5.28  1.4e-07 ***
## Year2007          0.3863      0.0723      5.34  9.8e-08 ***
## Year2008          0.1707      0.0790      2.16  0.03070 *
## Year2009          0.3410      0.0677      5.04  4.9e-07 ***
## Year2010          0.2090      0.0740      2.82  0.00479 **
## Year2011          0.2064      0.0738      2.80  0.00519 **
## Year2012          0.4045      0.0685      5.91  3.9e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.779
## Multiple R-squared:  0.0482, Adjusted R-squared:  0.043
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 486 is an outlier with |weight| = 0 ( < 3.2e-05);
## 255 weights are ~= 1. The remaining 2848 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.037  0.861  0.949   0.912  0.986  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.22e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016  1           1.008
## Year              1.016 16           1.001

```





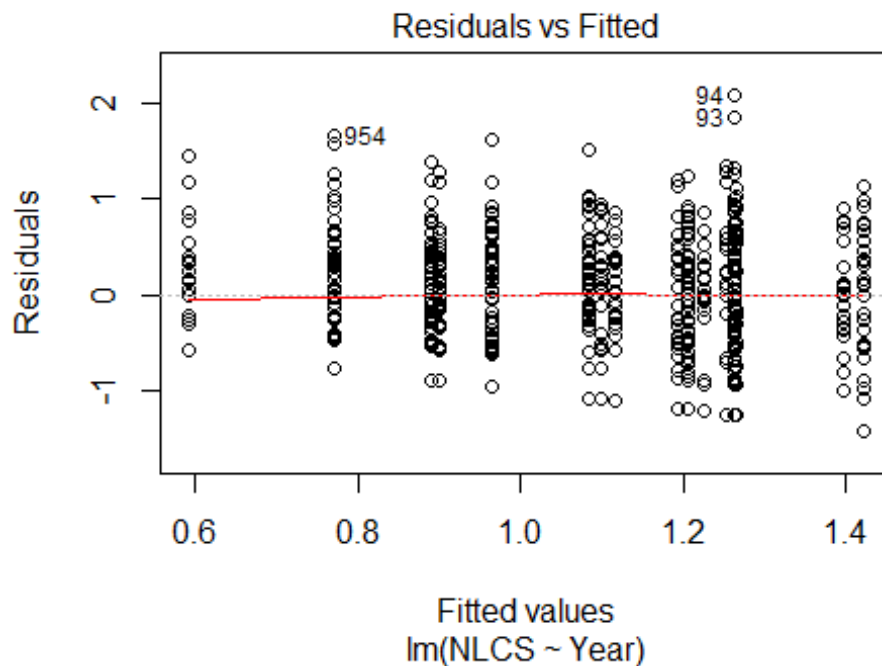
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1063 0007244732 3.574 1996    2200      3    2.596
## 1968 0031119324 4.853 1997    1210      2    3.862
## 2878 0031599055 3.579 1998    1405      2    2.684
## 4328 0346041912 3.749 2000    2200      1    2.555
## 5052 0036805915 3.818 2002    1210      2    2.541
## 5267 0036565284 4.562 2002    1210      2    3.285
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5257 -0.5501  0.0243  0.5276  3.8477
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9858     0.0485  20.33  < 2e-16 ***
## LastAuthorFemale1 0.0898     0.0423   2.12  0.03378 *
## Year1997         0.0194     0.0805   0.24  0.80912
## Year1998        -0.0792     0.0837  -0.95  0.34458
## Year1999         0.1466     0.0767   1.91  0.05601 .
## Year2000         0.2193     0.0748   2.93  0.00340 **
## Year2001         0.3482     0.0952   3.66  0.00026 ***
```

```

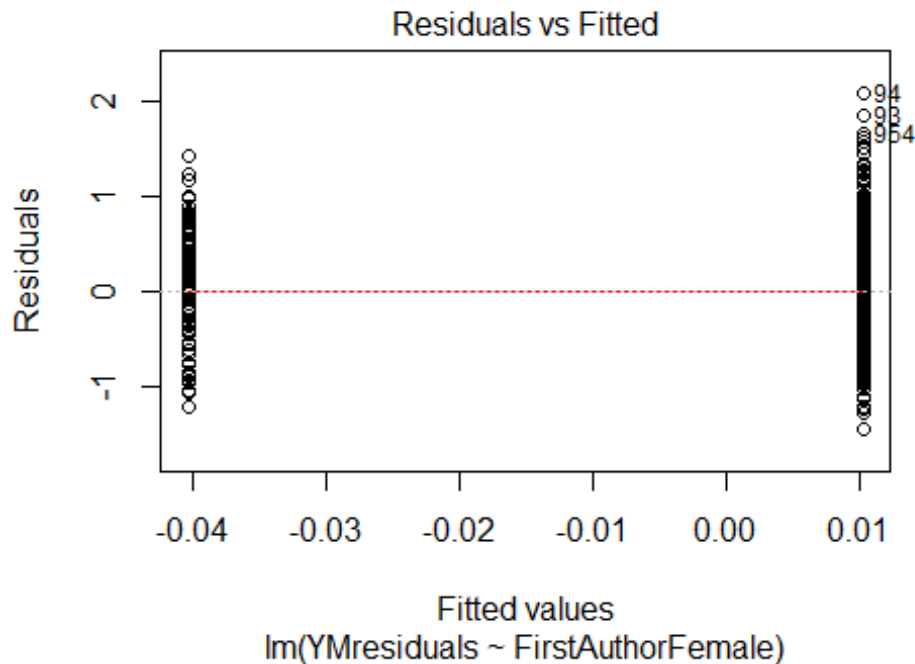
## Year2002          0.2971      0.0908      3.27  0.00107 **
## Year2003          0.3896      0.0761      5.12  3.2e-07 ***
## Year2004          0.4501      0.0748      6.02  1.9e-09 ***
## Year2005          0.4167      0.0778      5.36  9.2e-08 ***
## Year2006          0.3894      0.0716      5.44  5.9e-08 ***
## Year2007          0.3906      0.0726      5.38  8.1e-08 ***
## Year2008          0.1777      0.0795      2.23  0.02550 *
## Year2009          0.3494      0.0676      5.17  2.5e-07 ***
## Year2010          0.2204      0.0737      2.99  0.00280 **
## Year2011          0.2116      0.0740      2.86  0.00427 **
## Year2012          0.4084      0.0682      5.99  2.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.778
## Multiple R-squared:  0.046, Adjusted R-squared:  0.0408
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 486 is an outlier with |weight| = 0 ( < 3.2e-05);
## 262 weights are ~= 1. The remaining 2841 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0364 0.8600 0.9490 0.9120 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.22e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3104"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2201"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    52   56   66   66   86   73   89   68   79   62   88  108   79   57   91

```

```
## 2011 2012
## 103 89
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 23 27 24 21 32 20 47 20 30 27 50 66 42 31 55
## 2011 2012
## 73 65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 24 21 19 29 20 43 18 28 21 44 61 40 27 48
## 2011 2012
## 67 58
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.1
```

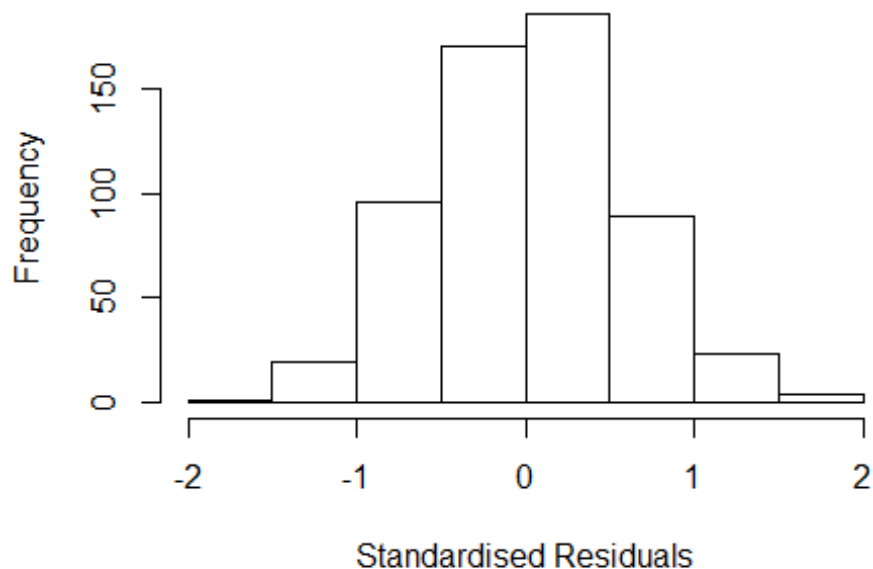


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.3, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.32071480923674"
## [1] "Male first author team size 2018 geometric mean: 2.2936847567063"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 580, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.94653971032285"
## [1] "Male last author team size 2018 geometric mean: 2.37834632589999"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 370, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.341 1      1.158
## LastAuthorFemale  1.339 1      1.157
## UniqueAuthors    1.537 4      1.055
## Year              1.743 16     1.018
```

## Residuals from first and last author and team size



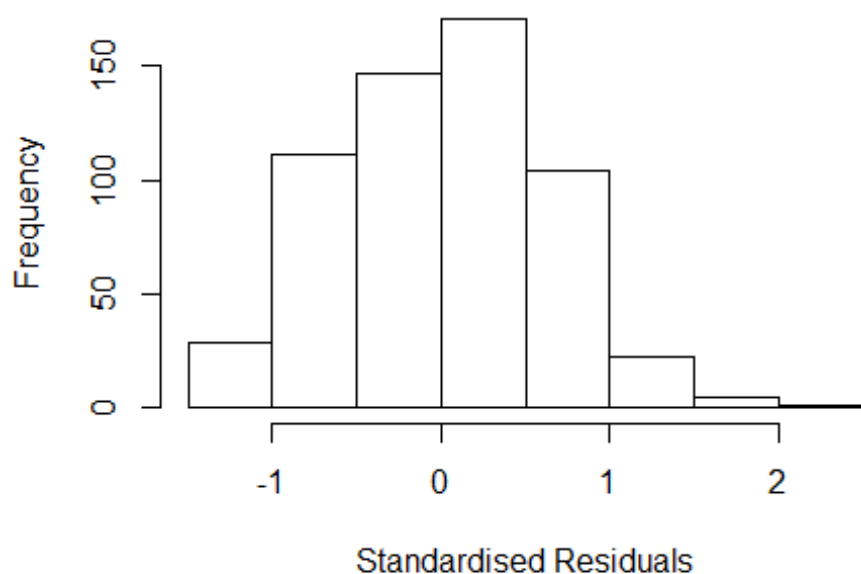
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6577 -0.4179 0.0137 0.4135 1.8935
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86628 0.17311 5.00 7.5e-07 ***
## FirstAuthorFemale1 -0.11797 0.06787 -1.74 0.083 .
## LastAuthorFemale1 0.02991 0.07095 0.42 0.673
## UniqueAuthors2 0.27725 0.06866 4.04 6.1e-05 ***
## UniqueAuthors3 0.52682 0.07067 7.45 3.4e-13 ***
## UniqueAuthors4 0.54736 0.09347 5.86 8.1e-09 ***
## UniqueAuthors5 0.67863 0.09104 7.45 3.4e-13 ***
## Year1997 0.11996 0.26671 0.45 0.653
## Year1998 0.06901 0.22269 0.31 0.757
## Year1999 0.09258 0.21019 0.44 0.660
```

```

## Year2000          0.19645    0.21262    0.92    0.356
## Year2001          0.13945    0.25510    0.55    0.585
## Year2002          0.11282    0.19377    0.58    0.561
## Year2003          0.38187    0.21467    1.78    0.076 .
## Year2004         -0.44834    0.19167   -2.34    0.020 *
## Year2005          0.05033    0.24068    0.21    0.834
## Year2006         -0.08679    0.20615   -0.42    0.674
## Year2007         -0.37906    0.19140   -1.98    0.048 *
## Year2008         -0.29034    0.19423   -1.49    0.136
## Year2009         -0.08505    0.19892   -0.43    0.669
## Year2010         -0.00287    0.19442   -0.01    0.988
## Year2011         -0.17483    0.19054   -0.92    0.359
## Year2012         -0.23207    0.18874   -1.23    0.219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.598
## Multiple R-squared:  0.255, Adjusted R-squared:  0.226
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 530 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.884  0.948  0.912  0.980  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.70e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.334 1      1.155
## LastAuthorFemale  1.360 1      1.166
## Year              1.143 16      1.004

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4093 -0.4649 0.0245 0.4433 2.1604
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12277 0.17134 6.55 1.3e-10 ***
## FirstAuthorFemale1 -0.06734 0.07293 -0.92 0.3562
## LastAuthorFemale1 0.02695 0.07637 0.35 0.7243
## Year1997 0.06082 0.28145 0.22 0.8290
## Year1998 -0.01158 0.23990 -0.05 0.9615
## Year1999 0.08363 0.20400 0.41 0.6820
## Year2000 0.28657 0.21590 1.33 0.1849
## Year2001 0.13814 0.27085 0.51 0.6102
## Year2002 0.15789 0.19959 0.79 0.4292
## Year2003 0.34162 0.21098 1.62 0.1060
## Year2004 -0.61635 0.19467 -3.17 0.0016 **
## Year2005 0.05872 0.22996 0.26 0.7986
```

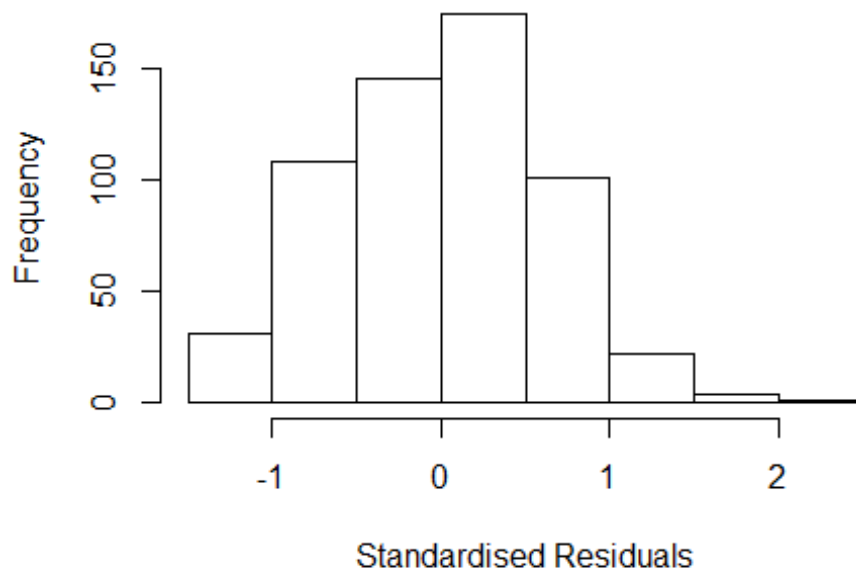
```

## Year2006          0.01110      0.20350      0.05      0.9565
## Year2007          -0.42018      0.19328      -2.17      0.0301 *
## Year2008          -0.23342      0.19243      -1.21      0.2256
## Year2009           0.00829      0.19983      0.04      0.9669
## Year2010           0.13115      0.19047      0.69      0.4914
## Year2011          -0.12580      0.18973      -0.66      0.5075
## Year2012          -0.22459      0.19228      -1.17      0.2433
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.668
## Multiple R-squared:  0.122, Adjusted R-squared:  0.0944
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 545 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.274  0.885   0.951   0.919   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.70e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.069 1      1.034
## Year              1.069 16      1.002

```



## Residuals from first author



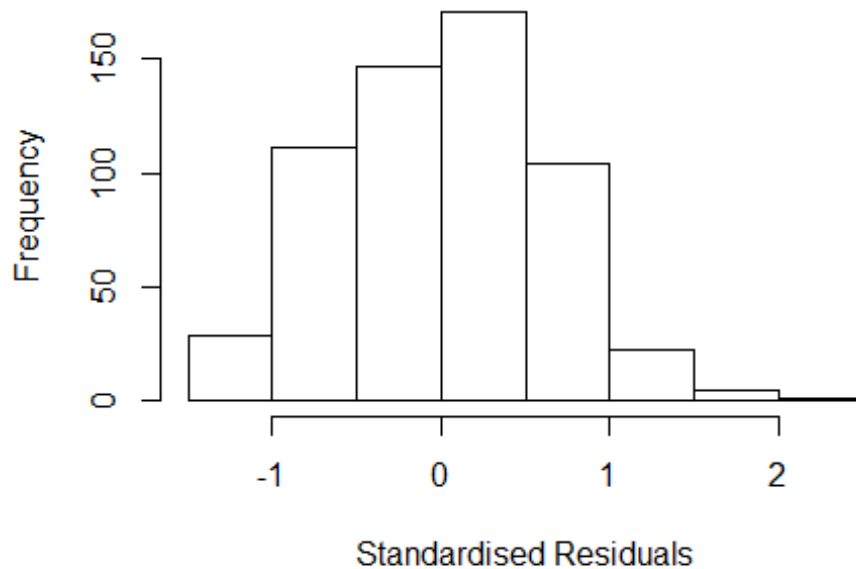
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4093 -0.4571 0.0254 0.4404 2.1584
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12753 0.17112 6.59 1e-10 ***
## FirstAuthorFemale1 -0.05649 0.06511 -0.87 0.3860
## Year1997 0.05803 0.28025 0.21 0.8360
## Year1998 -0.01363 0.23925 -0.06 0.9546
## Year1999 0.08245 0.20390 0.40 0.6861
## Year2000 0.28181 0.21582 1.31 0.1922
## Year2001 0.13924 0.26875 0.52 0.6046
## Year2002 0.15573 0.19960 0.78 0.4356
## Year2003 0.33795 0.21092 1.60 0.1096
## Year2004 -0.61707 0.19458 -3.17 0.0016 **
## Year2005 0.05540 0.22985 0.24 0.8096
## Year2006 0.00768 0.20346 0.04 0.9699
```

```

## Year2007          -0.41996    0.19308   -2.18    0.0300 *
## Year2008          -0.23557    0.19217   -1.23    0.2208
## Year2009           0.00641    0.19958    0.03    0.9744
## Year2010           0.12867    0.19027    0.68    0.4991
## Year2011          -0.12562    0.18960   -0.66    0.5079
## Year2012          -0.22830    0.19186   -1.19    0.2346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.675
## Multiple R-squared:  0.122, Adjusted R-squared:  0.0954
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 545 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.285  0.887   0.953   0.921   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.70e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.085 1          1.042
## Year              1.085 16          1.003

```

## Residuals from last author



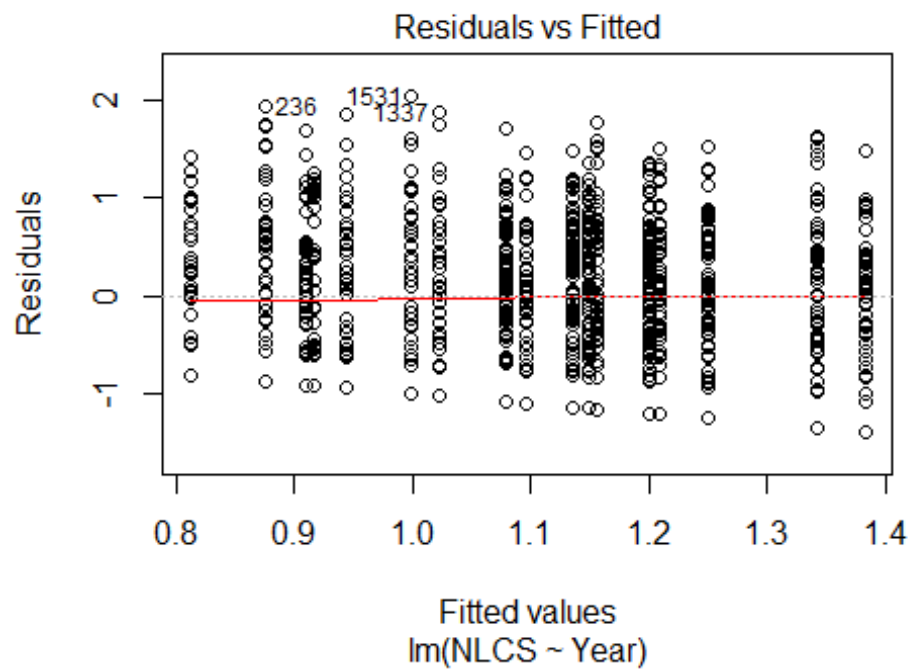
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4027 -0.4837 0.0342 0.4291 2.1659
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.123309 0.171741 6.54 1.4e-10 ***
## LastAuthorFemale1 -0.002264 0.068826 -0.03 0.9738
## Year1997 0.054759 0.282642 0.19 0.8464
## Year1998 -0.022101 0.238298 -0.09 0.9261
## Year1999 0.084380 0.204612 0.41 0.6802
## Year2000 0.279429 0.215653 1.30 0.1956
## Year2001 0.134865 0.271367 0.50 0.6194
## Year2002 0.155318 0.199702 0.78 0.4370
## Year2003 0.338602 0.211088 1.60 0.1092
## Year2004 -0.629538 0.193891 -3.25 0.0012 **
## Year2005 0.043969 0.230319 0.19 0.8487
## Year2006 0.000303 0.203056 0.00 0.9988
```

```

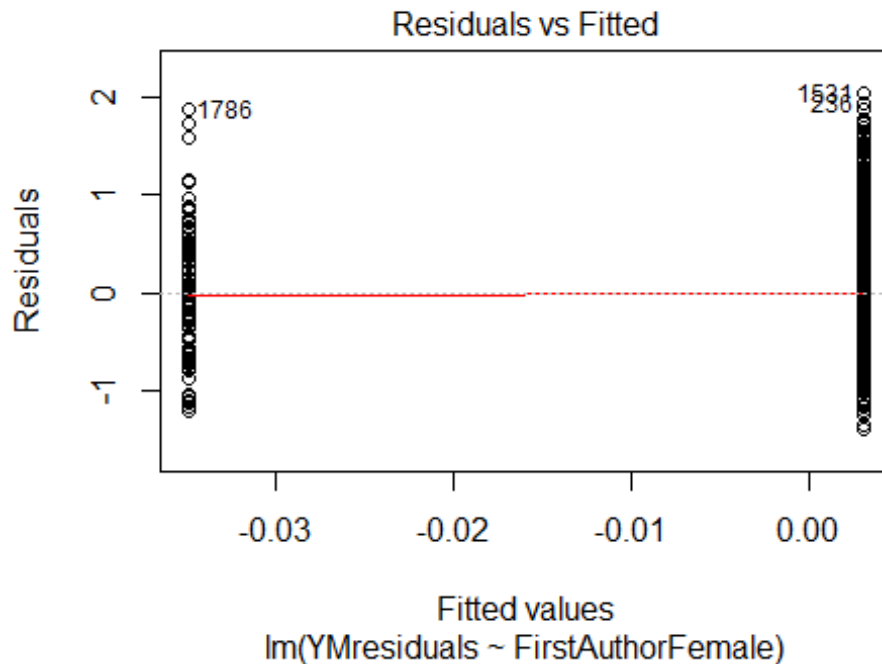
## Year2007          -0.430788    0.192789    -2.23    0.0258 *
## Year2008          -0.239708    0.192557    -1.24    0.2137
## Year2009           0.000177    0.199443     0.00    0.9993
## Year2010           0.118762    0.189481     0.63    0.5311
## Year2011          -0.135952    0.189312    -0.72    0.4730
## Year2012          -0.240700    0.191303    -1.26    0.2088
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.667
## Multiple R-squared:  0.121, Adjusted R-squared:  0.0949
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 547 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.270  0.888  0.951  0.919  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.70e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 588"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2202"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 182 243 206 231 194 218 130 160 167 167 239 217 176 266 253
## 2011 2012
## 256 243
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 69 49 61 60 44 53 49 55 57 76 75 65 109 100
## 2011 2012

```

```
## 108 111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 67 44 58 53 41 50 44 49 47 69 63 58 98 88
## 2011 2012
## 93 101
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```

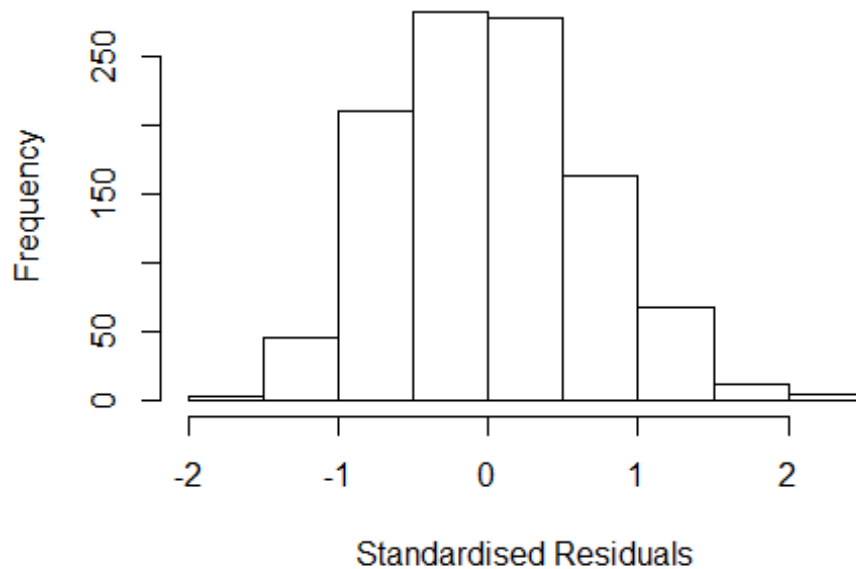


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.39, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 2.9953451699808"
## [1] "Male first author team size 2018 geometric mean: 2.65416065373692"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 690, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.47939698673123"
## [1] "Male last author team size 2018 geometric mean: 2.72876469804396"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.340 1          1.158
## LastAuthorFemale  1.362 1          1.167
## UniqueAuthors    1.436 4          1.046
## Year              1.535 16         1.013
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6159 -0.4947 -0.0124 0.4751 2.4129
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4947 0.1031 4.80 1.8e-06 ***
## FirstAuthorFemale1 -0.1141 0.0821 -1.39 0.16476
## LastAuthorFemale1 0.0598 0.0852 0.70 0.48284
## UniqueAuthors2 0.4945 0.0543 9.11 < 2e-16 ***
## UniqueAuthors3 0.5996 0.0656 9.15 < 2e-16 ***
## UniqueAuthors4 0.6804 0.0814 8.35 < 2e-16 ***
## UniqueAuthors5 0.4763 0.1119 4.26 2.3e-05 ***
## Year1997 0.0955 0.1403 0.68 0.49651
## Year1998 0.1368 0.1607 0.85 0.39482
## Year1999 0.1503 0.1429 1.05 0.29330
```

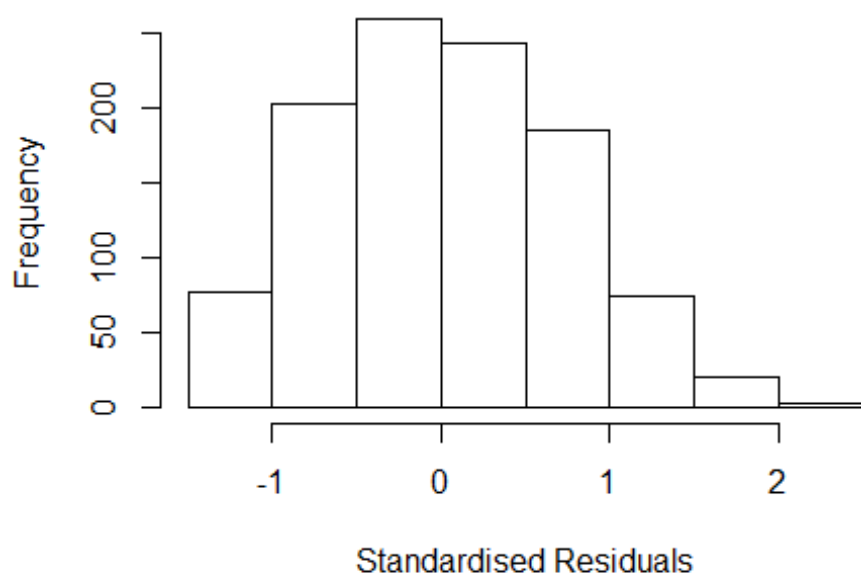
```

## Year2000          0.4660      0.1548      3.01  0.00268 **
## Year2001          0.1456      0.1396      1.04  0.29704
## Year2002          0.1283      0.1336      0.96  0.33707
## Year2003          0.2229      0.1550      1.44  0.15056
## Year2004          0.2488      0.1322      1.88  0.06018 .
## Year2005          0.5216      0.1312      3.97  7.5e-05 ***
## Year2006          0.3043      0.1410      2.16  0.03115 *
## Year2007          0.4294      0.1300      3.30  0.00099 ***
## Year2008          0.3527      0.1360      2.59  0.00961 **
## Year2009          0.3229      0.1170      2.76  0.00590 **
## Year2010          0.2743      0.1225      2.24  0.02542 *
## Year2011          0.2080      0.1183      1.76  0.07896 .
## Year2012          0.1821      0.1180      1.54  0.12295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.659
## Multiple R-squared:  0.188, Adjusted R-squared:  0.171
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 976 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.880  0.943  0.907  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.201 1      1.096
## LastAuthorFemale  1.186 1      1.089
## Year              1.079 16      1.002

```



## Residuals from first and last author



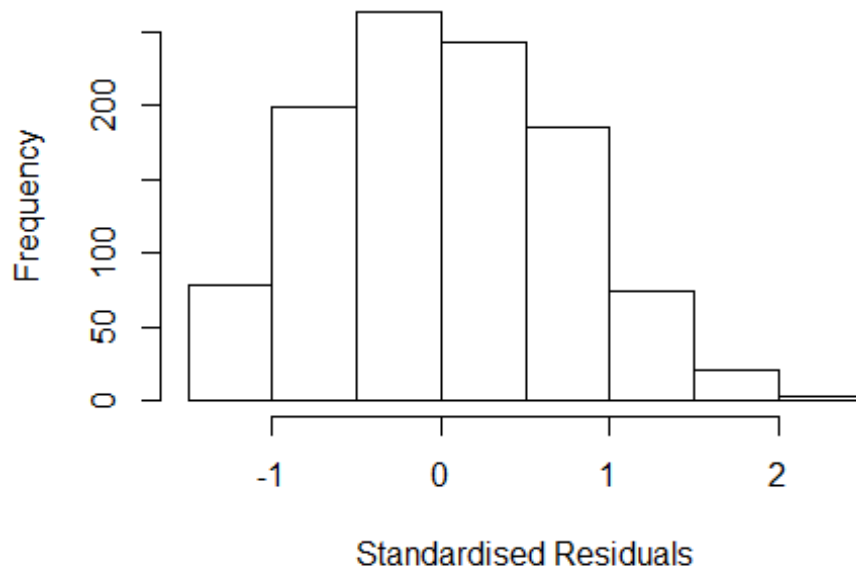
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4035 -0.5198 -0.0322 0.5324 2.1582
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7348 0.1055 6.97 5.8e-12 ***
## FirstAuthorFemale1 -0.0580 0.0878 -0.66 0.50866
## LastAuthorFemale1 0.0453 0.0933 0.49 0.62764
## Year1997 0.0609 0.1506 0.40 0.68594
## Year1998 0.0702 0.1675 0.42 0.67517
## Year1999 0.1280 0.1485 0.86 0.38909
## Year2000 0.5360 0.1542 3.48 0.00053 ***
## Year2001 0.1543 0.1558 0.99 0.32213
## Year2002 0.1430 0.1583 0.90 0.36662
## Year2003 0.1290 0.1608 0.80 0.42254
## Year2004 0.2768 0.1430 1.94 0.05325 .
## Year2005 0.6687 0.1428 4.68 3.2e-06 ***
```

```

## Year2006          0.4192      0.1428      2.93  0.00341 **
## Year2007          0.5194      0.1413      3.68  0.00025 ***
## Year2008          0.4791      0.1469      3.26  0.00115 **
## Year2009          0.4349      0.1216      3.58  0.00036 ***
## Year2010          0.3720      0.1303      2.85  0.00440 **
## Year2011          0.3456      0.1285      2.69  0.00728 **
## Year2012          0.3168      0.1261      2.51  0.01216 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.748
## Multiple R-squared:  0.0546, Adjusted R-squared:  0.0384
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 1003 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.386  0.883   0.950   0.919   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1          1.023
## Year              1.047 16          1.001

```

## Residuals from first author



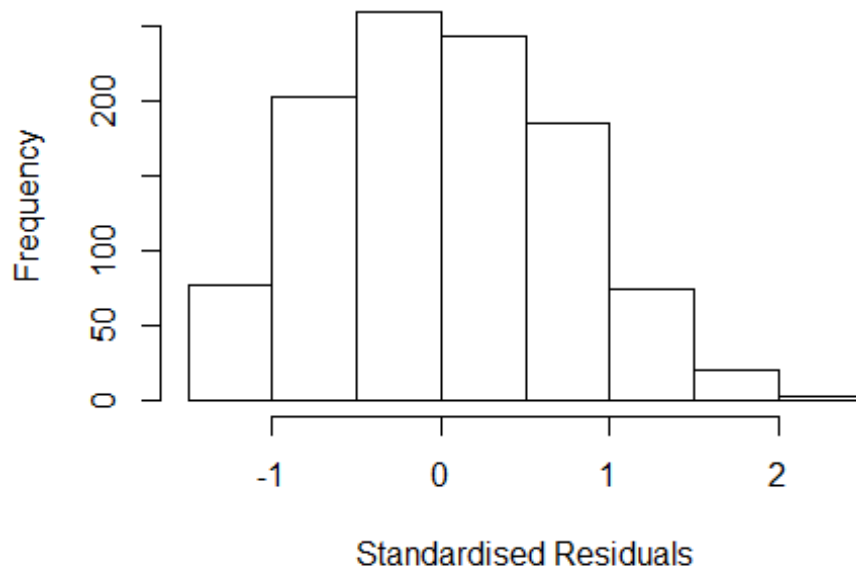
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4057 -0.5197 -0.0294 0.5320 2.1577
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7373 0.1053 7.00 4.5e-12 ***
## FirstAuthorFemale1 -0.0448 0.0810 -0.55 0.58025
## Year1997 0.0610 0.1505 0.41 0.68519
## Year1998 0.0680 0.1677 0.41 0.68534
## Year1999 0.1279 0.1487 0.86 0.38987
## Year2000 0.5329 0.1540 3.46 0.00056 ***
## Year2001 0.1531 0.1557 0.98 0.32550
## Year2002 0.1410 0.1581 0.89 0.37246
## Year2003 0.1280 0.1607 0.80 0.42595
## Year2004 0.2772 0.1429 1.94 0.05270 .
## Year2005 0.6684 0.1429 4.68 3.3e-06 ***
## Year2006 0.4185 0.1429 2.93 0.00348 **
```

```

## Year2007          0.5206      0.1416      3.68  0.00025 ***
## Year2008          0.4782      0.1469      3.26  0.00117 **
## Year2009          0.4325      0.1213      3.56  0.00038 ***
## Year2010          0.3721      0.1302      2.86  0.00435 **
## Year2011          0.3471      0.1285      2.70  0.00703 **
## Year2012          0.3163      0.1261      2.51  0.01231 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.749
## Multiple R-squared:  0.0543, Adjusted R-squared:  0.039
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 1003 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.387  0.883  0.950  0.919  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1      1.016
## Year      1.032 16      1.001

```

## Residuals from last author



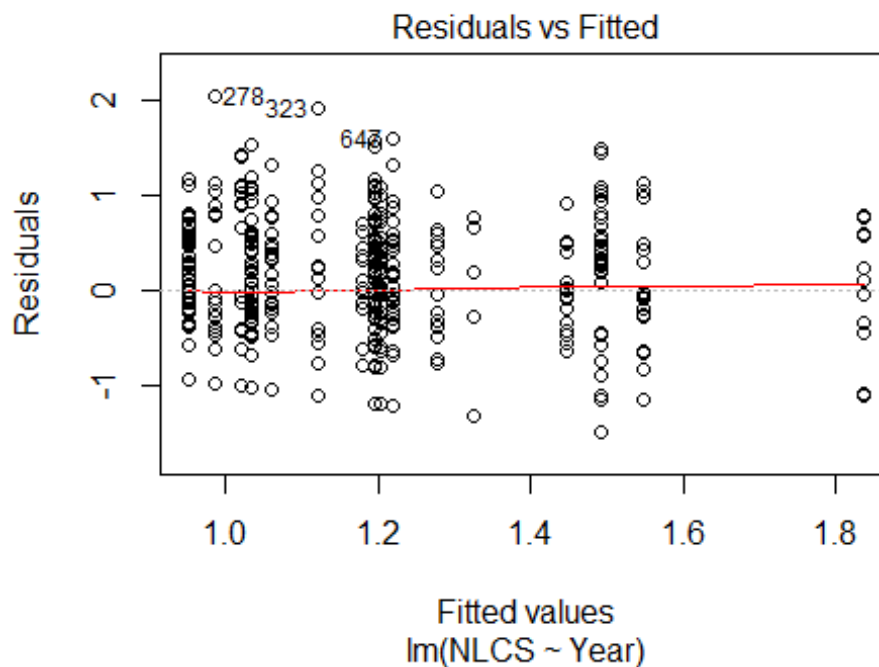
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3986 -0.5180 -0.0316 0.5289 2.1603
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7319 0.1054 6.94 6.8e-12 ***
## LastAuthorFemale1 0.0247 0.0892 0.28 0.78173
## Year1997 0.0632 0.1505 0.42 0.67482
## Year1998 0.0693 0.1678 0.41 0.67963
## Year1999 0.1301 0.1486 0.87 0.38179
## Year2000 0.5326 0.1539 3.46 0.00056 ***
## Year2001 0.1572 0.1552 1.01 0.31153
## Year2002 0.1438 0.1582 0.91 0.36359
## Year2003 0.1333 0.1602 0.83 0.40580
## Year2004 0.2734 0.1430 1.91 0.05620 .
## Year2005 0.6667 0.1429 4.67 3.5e-06 ***
## Year2006 0.4201 0.1431 2.94 0.00340 **
```

```

## Year2007          0.5191      0.1413      3.67  0.00025 ***
## Year2008          0.4772      0.1475      3.24  0.00125 **
## Year2009          0.4340      0.1216      3.57  0.00037 ***
## Year2010          0.3700      0.1304      2.84  0.00464 **
## Year2011          0.3441      0.1287      2.67  0.00764 **
## Year2012          0.3153      0.1264      2.50  0.01275 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.749
## Multiple R-squared:  0.0541, Adjusted R-squared:  0.0388
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 1006 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.386  0.883  0.952  0.919  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1066"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2203"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   34   53   49   46   57   50   37   39   39   65   66   62  115   90
## 2011 2012
##   95  104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5   11   21   17   21   23   22   14   13   16   17   35   25   62   46
## 2011 2012

```

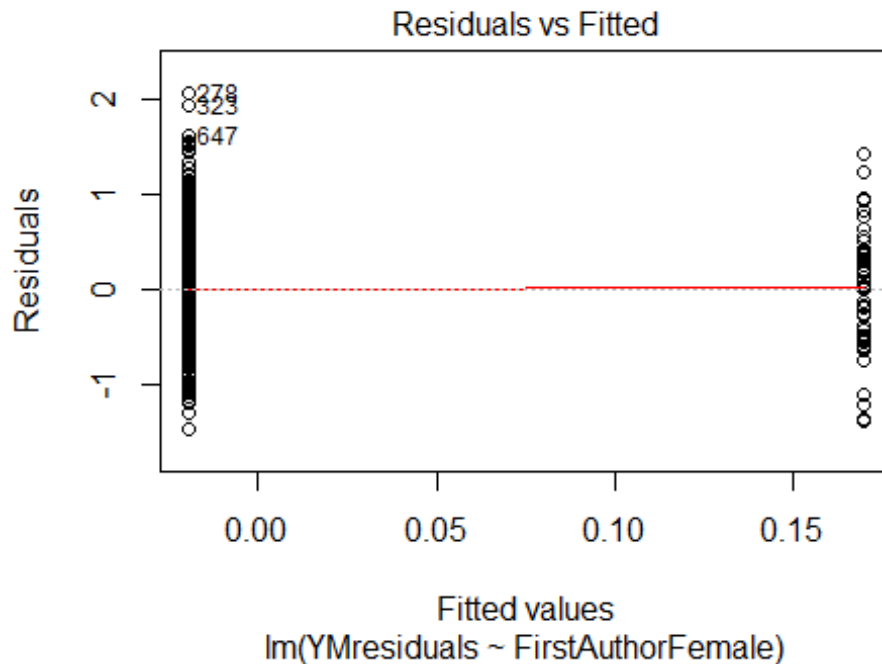
```
## 53 63
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 4 11 19 15 20 23 22 12 12 14 14 30 23 54 39
## 2011 2012
## 48 51
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.095, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 2.56260005280024"
## [1] "Male first author team size 2018 geometric mean: 2.60696004226789"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

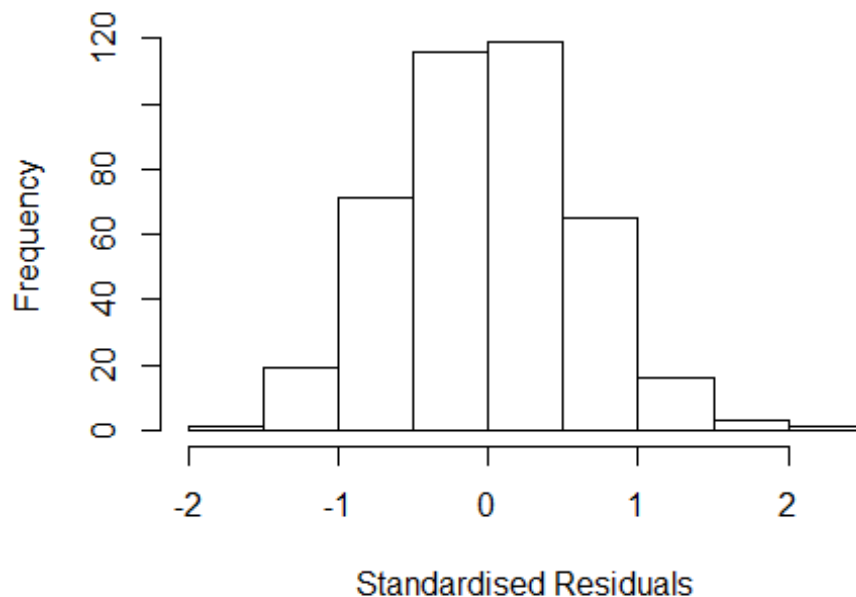


```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.14810629603756"
## [1] "Male last author team size 2018 geometric mean: 2.7218508535737"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 260, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.383	1	1.176
LastAuthorFemale	1.490	1	1.221
UniqueAuthors	2.010	4	1.091
Year	2.592	16	1.030



## Residuals from first and last author and team size



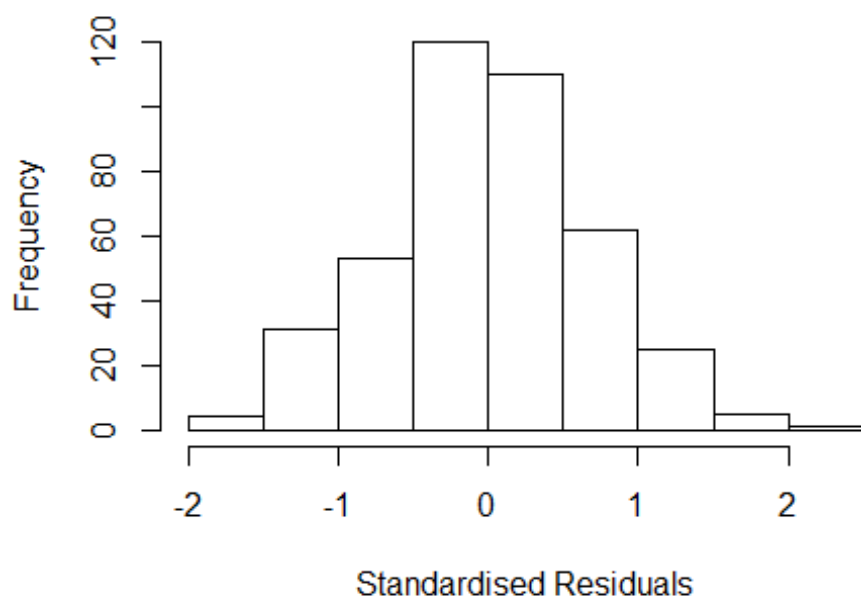
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.65243 -0.44309 -0.00713 0.43368 2.19543
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9506 0.3064 3.10 0.0021 **
## FirstAuthorFemale1 0.2250 0.1120 2.01 0.0453 *
## LastAuthorFemale1 -0.0020 0.1296 -0.02 0.9877
## UniqueAuthors2 0.4354 0.0935 4.66 4.4e-06 ***
## UniqueAuthors3 0.4517 0.0958 4.72 3.4e-06 ***
## UniqueAuthors4 0.5728 0.1236 4.64 4.9e-06 ***
## UniqueAuthors5 0.3416 0.1512 2.26 0.0244 *
## Year1997 0.7764 0.3951 1.96 0.0501 .
## Year1998 -0.3055 0.3833 -0.80 0.4260
## Year1999 0.2117 0.3251 0.65 0.5153
```

```

## Year2000          0.1785      0.3616      0.49      0.6218
## Year2001         -0.3442      0.3362     -1.02      0.3066
## Year2002         -0.1100      0.3441     -0.32      0.7494
## Year2003          0.2665      0.4604      0.58      0.5630
## Year2004         -0.0960      0.3372     -0.28      0.7760
## Year2005          0.0388      0.3248      0.12      0.9049
## Year2006          0.3057      0.3411      0.90      0.3707
## Year2007         -0.0111      0.3207     -0.03      0.9724
## Year2008         -0.2984      0.3352     -0.89      0.3739
## Year2009         -0.3024      0.3176     -0.95      0.3416
## Year2010         -0.1988      0.3154     -0.63      0.5288
## Year2011         -0.4044      0.3156     -1.28      0.2007
## Year2012         -0.0297      0.3181     -0.09      0.9257
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.639
## Multiple R-squared:  0.198, Adjusted R-squared:  0.153
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 386 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.214  0.873   0.953   0.910   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.273 1      1.128
## LastAuthorFemale  1.358 1      1.165
## Year              1.363 16      1.010

```

## Residuals from first and last author



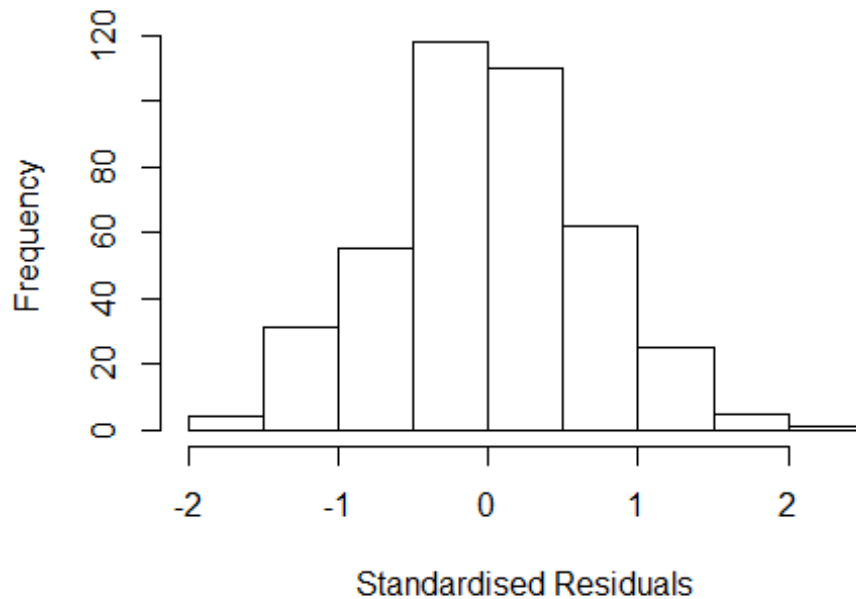
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.690 -0.431 -0.013 0.455 2.187
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19010 0.41985 2.83 0.0048 **
## FirstAuthorFemale1 0.23884 0.11497 2.08 0.0384 *
## LastAuthorFemale1 -0.01048 0.13469 -0.08 0.9380
## Year1997 0.68049 0.47247 1.44 0.1506
## Year1998 -0.37554 0.49644 -0.76 0.4498
## Year1999 0.19192 0.43482 0.44 0.6592
## Year2000 0.22677 0.46859 0.48 0.6287
## Year2001 -0.35199 0.45320 -0.78 0.4378
## Year2002 -0.14849 0.47479 -0.31 0.7546
## Year2003 0.50015 0.59286 0.84 0.3994
## Year2004 0.00348 0.44730 0.01 0.9938
## Year2005 0.16341 0.44517 0.37 0.7138
```

```

## Year2006          0.35251    0.45471    0.78    0.4387
## Year2007          0.10323    0.43285    0.24    0.8116
## Year2008         -0.17052    0.44556   -0.38    0.7022
## Year2009         -0.18598    0.42935   -0.43    0.6651
## Year2010         -0.06879    0.42813   -0.16    0.8724
## Year2011         -0.28556    0.42968   -0.66    0.5067
## Year2012          0.04478    0.43214    0.10    0.9175
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.656
## Multiple R-squared:  0.12,   Adjusted R-squared:  0.0796
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 381 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.857  0.952  0.905  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.43e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.126 1      1.061
## Year              1.126 16      1.004

```

## Residuals from first author



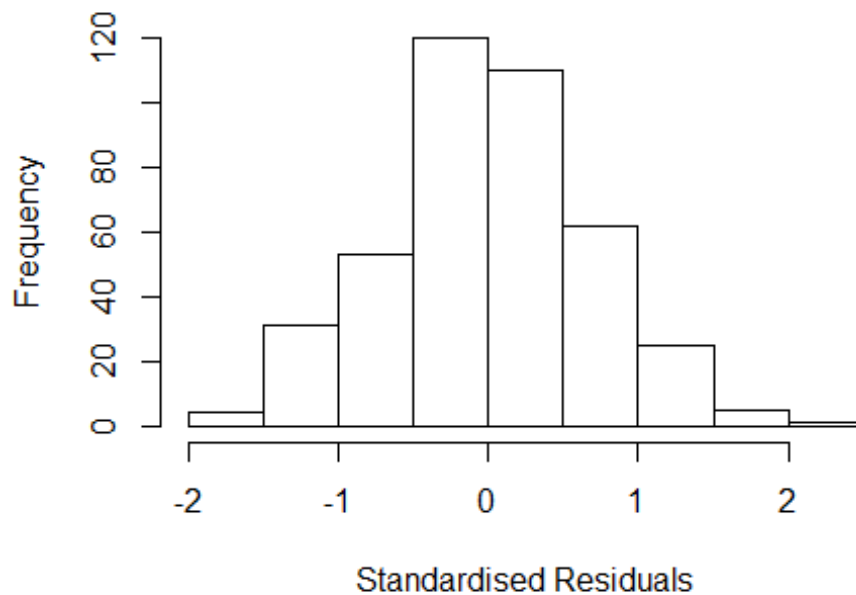
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6926 -0.4289 -0.0124 0.4550 2.1879
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19060 0.42025 2.83 0.0048 **
## FirstAuthorFemale1 0.23547 0.10837 2.17 0.0304 *
## Year1997 0.68028 0.47289 1.44 0.1511
## Year1998 -0.37712 0.49718 -0.76 0.4486
## Year1999 0.19083 0.43511 0.44 0.6612
## Year2000 0.22495 0.46855 0.48 0.6314
## Year2001 -0.35352 0.45361 -0.78 0.4362
## Year2002 -0.14949 0.47516 -0.31 0.7532
## Year2003 0.50205 0.59250 0.85 0.3973
## Year2004 0.00232 0.44761 0.01 0.9959
## Year2005 0.16076 0.44498 0.36 0.7181
## Year2006 0.34893 0.45183 0.77 0.4404
```

```

## Year2007          0.10255    0.43326    0.24    0.8130
## Year2008          -0.17109    0.44598   -0.38    0.7015
## Year2009          -0.18668    0.42975   -0.43    0.6642
## Year2010          -0.07057    0.42821   -0.16    0.8692
## Year2011          -0.28670    0.43003   -0.67    0.5054
## Year2012           0.04386    0.43257    0.10    0.9193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.653
## Multiple R-squared:  0.12,   Adjusted R-squared:  0.0823
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 381 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.239  0.856  0.952  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.202 1          1.097
## Year            1.202 16          1.006

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7084 -0.4151 -0.0154 0.4642 2.1358
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18902 0.41901 2.84 0.0048 **
## LastAuthorFemale1 0.07651 0.12364 0.62 0.5364
## Year1997 0.68094 0.47157 1.44 0.1495
## Year1998 -0.37494 0.49860 -0.75 0.4525
## Year1999 0.20384 0.43464 0.47 0.6393
## Year2000 0.26882 0.47145 0.57 0.5689
## Year2001 -0.29978 0.45118 -0.66 0.5068
## Year2002 -0.13042 0.47333 -0.28 0.7830
## Year2003 0.51937 0.59466 0.87 0.3830
## Year2004 -0.00196 0.44538 0.00 0.9965
## Year2005 0.14689 0.44388 0.33 0.7409
## Year2006 0.36405 0.45286 0.80 0.4220
```

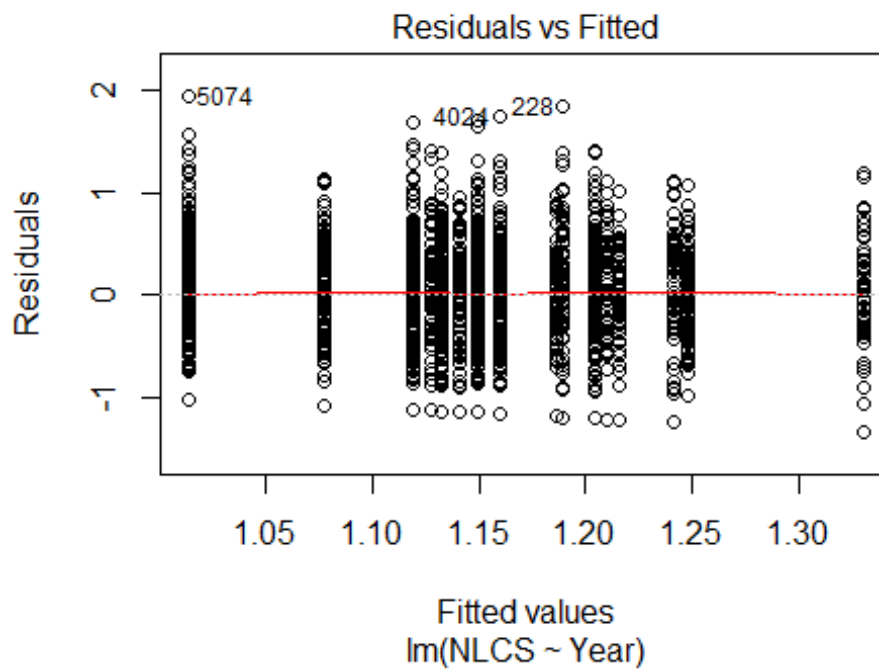
```

## Year2007      0.13653      0.43191      0.32      0.7521
## Year2008     -0.13443      0.44292     -0.30      0.7617
## Year2009     -0.16493      0.42824     -0.39      0.7003
## Year2010     -0.07209      0.42748     -0.17      0.8662
## Year2011     -0.25921      0.42901     -0.60      0.5461
## Year2012      0.05769      0.43174      0.13      0.8938
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.112, Adjusted R-squared:  0.0733
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 380 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.274  0.854  0.952  0.905  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.43e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 411"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2204"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  195  191  179  177  202  222  197  172  183  247  242  329  343  427  423
## 2011 2012
##  466  487
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   77   64   52   66   63   76   85   59  103  102  137  175  193  220  229
## 2011 2012

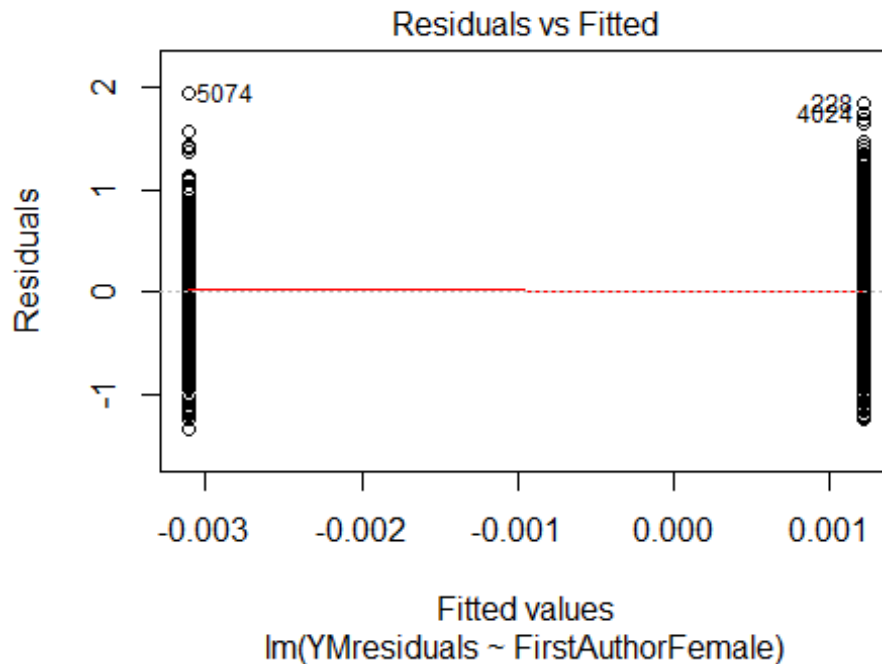
```



```
## 288 276
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 60 40 57 54 69 74 51 87 87 115 151 168 196 203
## 2011 2012
## 254 238
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 72, df = 16, p-value = 5e-09
```

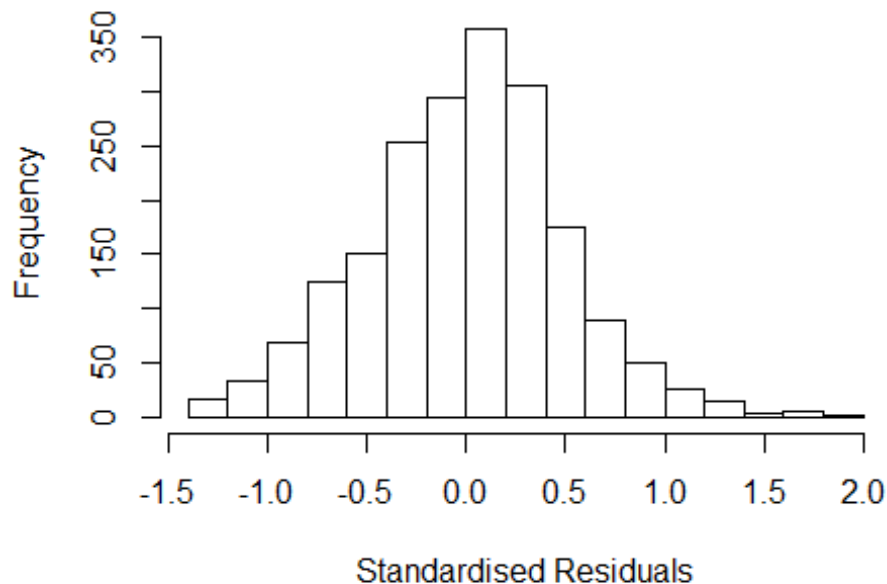


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.15369525673546"
## [1] "Male first author team size 2018 geometric mean: 3.71763095180388"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.85846588986749"
## [1] "Male last author team size 2018 geometric mean: 3.84930482533808"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7400, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.060 1      1.030
## LastAuthorFemale  1.040 1      1.020
## UniqueAuthors    1.300 4      1.033
## Year             1.319 16      1.009
```

## Residuals from first and last author and team size



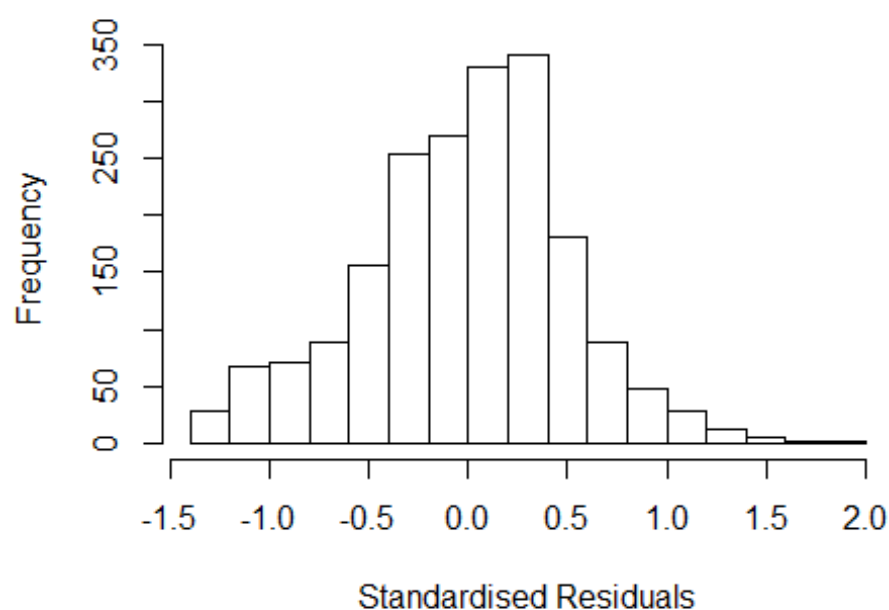
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3650 -0.3126  0.0269  0.3080  1.8579
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.95343    0.08804   10.83 < 2e-16 ***
## FirstAuthorFemale1 -0.01634    0.02467   -0.66  0.50778
## LastAuthorFemale1  0.01927    0.03078    0.63  0.53140
## UniqueAuthors2     0.33684    0.05052    6.67 3.4e-11 ***
## UniqueAuthors3     0.38972    0.04697    8.30 < 2e-16 ***
## UniqueAuthors4     0.37594    0.04793    7.84 7.2e-15 ***
## UniqueAuthors5     0.47143    0.04560   10.34 < 2e-16 ***
## Year1997          -0.00504    0.12040   -0.04  0.96661
## Year1998           0.09103    0.11793    0.77  0.44030
## Year1999          -0.10050    0.11139   -0.90  0.36704
```

```

## Year2000      -0.06255    0.12458   -0.50  0.61565
## Year2001      -0.17179    0.10495   -1.64  0.10183
## Year2002      -0.10077    0.10603   -0.95  0.34207
## Year2003      -0.04439    0.10279   -0.43  0.66589
## Year2004      -0.08719    0.09276   -0.94  0.34735
## Year2005      -0.11284    0.09619   -1.17  0.24093
## Year2006      -0.19976    0.09714   -2.06  0.03988 *
## Year2007      -0.18430    0.09742   -1.89  0.05867 .
## Year2008      -0.09276    0.09319   -1.00  0.31963
## Year2009      -0.15988    0.09204   -1.74  0.08254 .
## Year2010      -0.19952    0.08988   -2.22  0.02655 *
## Year2011      -0.31540    0.08991   -3.51  0.00046 ***
## Year2012      -0.19259    0.09019   -2.14  0.03284 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0977
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0684 0.8580 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.07e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## LastAuthorFemale 1.037 1          1.018
## Year 1.074 16          1.002

```

## Residuals from first and last author

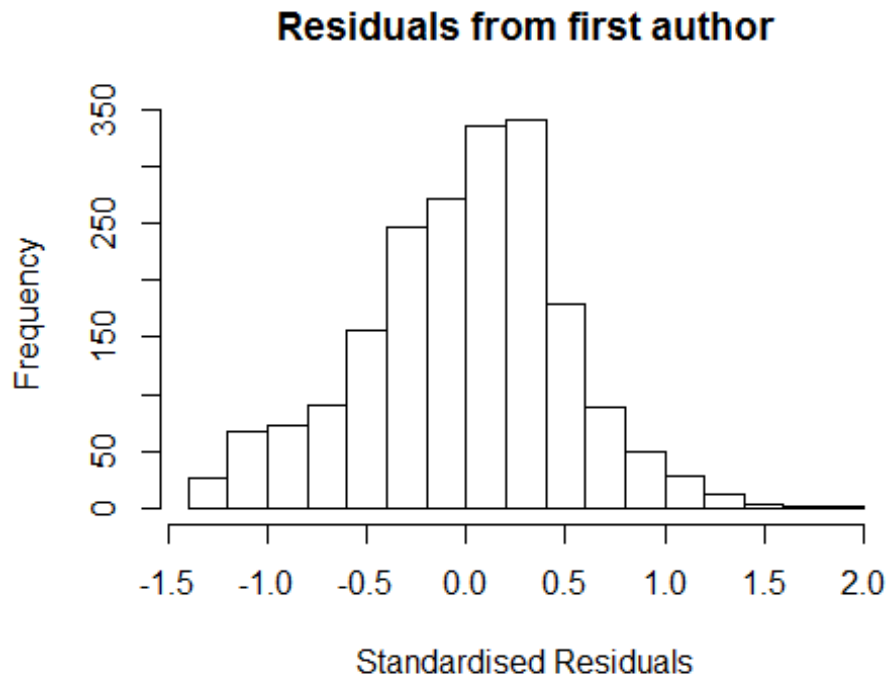


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3334 -0.3103  0.0357  0.3081  1.9511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1900     0.0957  12.44  <2e-16 ***
## FirstAuthorFemale1  0.0134     0.0252   0.53   0.594
## LastAuthorFemale1  0.0150     0.0325   0.46   0.644
## Year1997          0.0485     0.1369   0.35   0.723
## Year1998          0.1300     0.1315   0.99   0.323
## Year1999         -0.0467     0.1150  -0.41   0.685
## Year2000          0.0715     0.1309   0.55   0.585
## Year2001         -0.0643     0.1196  -0.54   0.591
## Year2002         -0.0170     0.1161  -0.15   0.884
## Year2003          0.0531     0.1118   0.48   0.635
## Year2004          0.0283     0.1036   0.27   0.785
## Year2005          0.0270     0.1045   0.26   0.796
```

```

## Year2006          -0.0997      0.1075   -0.93    0.354
## Year2007          -0.0681      0.1086   -0.63    0.531
## Year2008           0.0243      0.1024    0.24    0.812
## Year2009          -0.0415      0.1015   -0.41    0.682
## Year2010          -0.0650      0.1001   -0.65    0.516
## Year2011          -0.2035      0.1012   -2.01    0.044 *
## Year2012          -0.0690      0.1001   -0.69    0.490
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.0147
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1828 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0401 0.8590 0.9530 0.8900 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.035 1      1.017
## Year              1.035 16      1.001

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3375 -0.3087 0.0336 0.3130 1.9467
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1926 0.0950 12.56 <2e-16 ***
## FirstAuthorFemale1 0.0149 0.0254 0.59 0.559
## Year1997 0.0468 0.1366 0.34 0.732
## Year1998 0.1300 0.1313 0.99 0.322
## Year1999 -0.0486 0.1147 -0.42 0.672
## Year2000 0.0696 0.1303 0.53 0.594
## Year2001 -0.0646 0.1192 -0.54 0.588
## Year2002 -0.0181 0.1157 -0.16 0.876
## Year2003 0.0521 0.1114 0.47 0.640
## Year2004 0.0270 0.1032 0.26 0.794
## Year2005 0.0261 0.1041 0.25 0.802
## Year2006 -0.1005 0.1072 -0.94 0.348
```

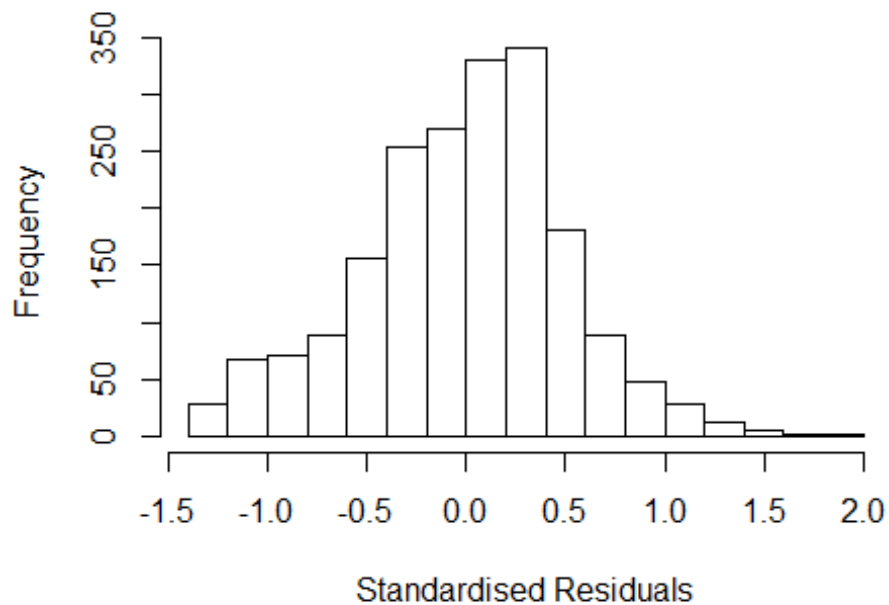
```

## Year2007          -0.0680      0.1084   -0.63    0.530
## Year2008           0.0238      0.1020    0.23    0.815
## Year2009          -0.0422      0.1011   -0.42    0.676
## Year2010          -0.0656      0.0998   -0.66    0.511
## Year2011          -0.2032      0.1009   -2.01    0.044 *
## Year2012          -0.0695      0.0998   -0.70    0.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.0151
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1826 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0426 0.8590 0.9540 0.8900 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.038 1          1.019
## Year            1.038 16          1.001

```



## Residuals from last author



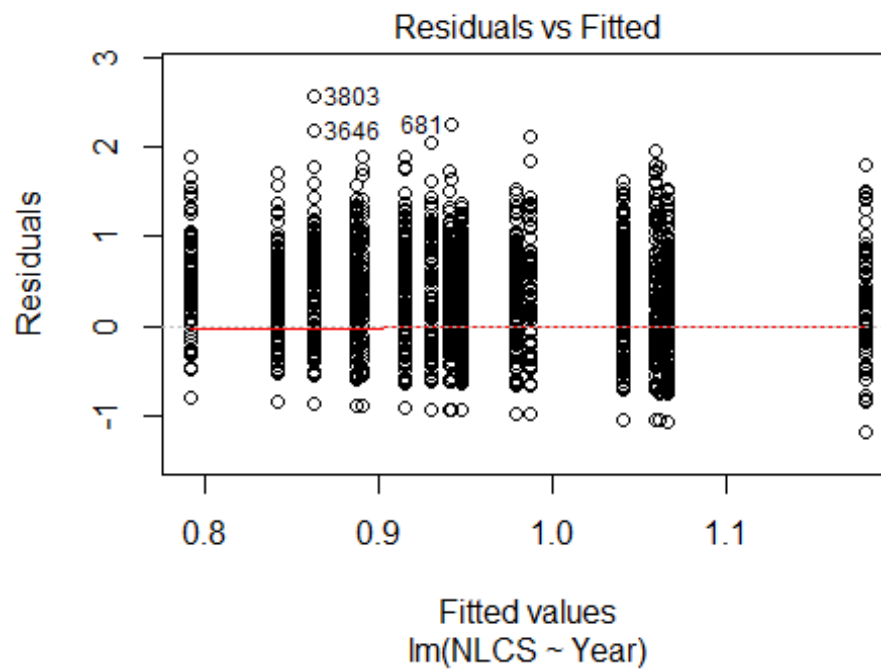
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3231 -0.3088 0.0384 0.3111 1.9612
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1919 0.0957 12.46 <2e-16 ***
## LastAuthorFemale1 0.0170 0.0327 0.52 0.604
## Year1997 0.0480 0.1372 0.35 0.727
## Year1998 0.1312 0.1315 1.00 0.318
## Year1999 -0.0473 0.1151 -0.41 0.682
## Year2000 0.0720 0.1312 0.55 0.583
## Year2001 -0.0630 0.1196 -0.53 0.599
## Year2002 -0.0156 0.1162 -0.13 0.893
## Year2003 0.0539 0.1119 0.48 0.630
## Year2004 0.0300 0.1037 0.29 0.772
## Year2005 0.0287 0.1046 0.27 0.784
## Year2006 -0.0978 0.1077 -0.91 0.364
```

```

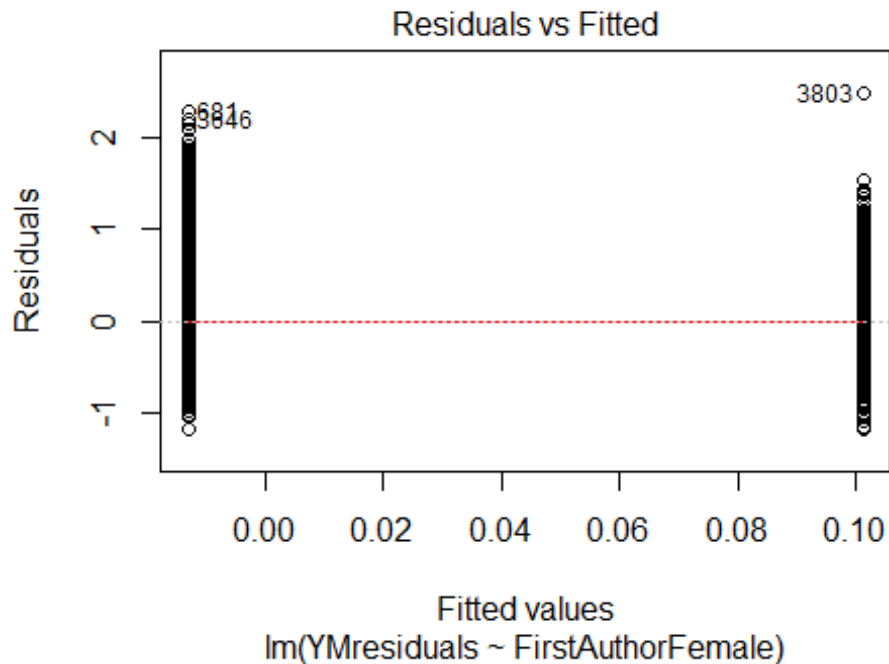
## Year2007          -0.0658      0.1086    -0.61      0.545
## Year2008           0.0262      0.1025      0.26      0.798
## Year2009          -0.0389      0.1015    -0.38      0.701
## Year2010          -0.0630      0.1002    -0.63      0.530
## Year2011          -0.2020      0.1013    -2.00      0.046 *
## Year2012          -0.0665      0.1001    -0.66      0.507
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.0151
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1832 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0366 0.8580 0.9530 0.8900 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1973"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2205"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 383 383 441 439 383 428 325 361 367 349 400 446 480 440 484
## 2011 2012
## 464 562
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 90 72 97 104 88 98 136 187 189 167 174 219 232 219 235
## 2011 2012

```

```
## 253 315
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 68 82 102 85 92 127 169 181 152 155 205 205 192 212
## 2011 2012
## 224 270
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 50, df = 16, p-value = 3e-05
```

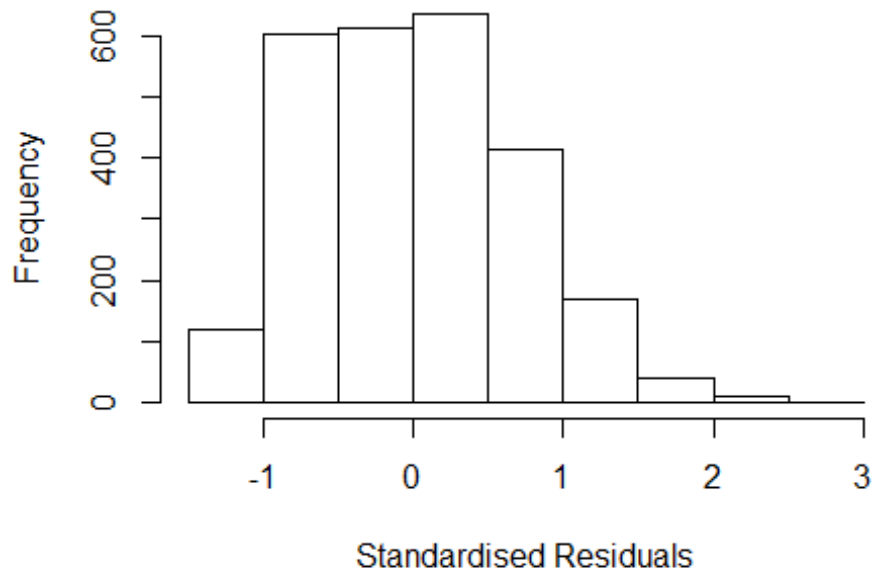


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.1, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 2.80555219181807"
## [1] "Male first author team size 2018 geometric mean: 2.58004897528783"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.42224688865847"
## [1] "Male last author team size 2018 geometric mean: 2.67444043681698"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7600, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.352 1          1.163
## LastAuthorFemale  1.342 1          1.159
## UniqueAuthors    1.160 4          1.019
## Year              1.168 16         1.005
```

## Residuals from first and last author and team size



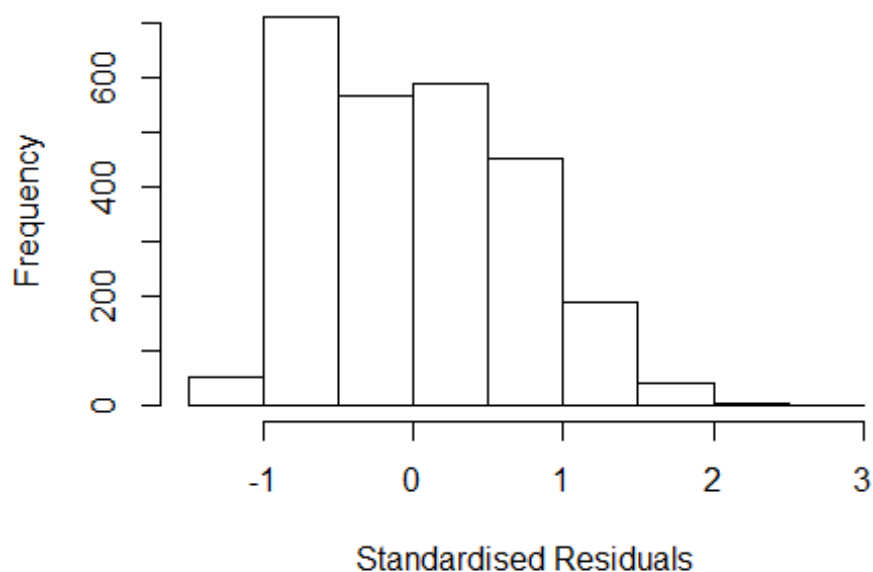
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3646 9444267658 3.043 2004      1400      4      2.529
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.491 -0.514 -0.020  0.486  2.529
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.60673    0.09826   6.17 7.7e-10 ***
## FirstAuthorFemale1 0.11817    0.04815   2.45  0.014 *
## LastAuthorFemale1 -0.07721    0.05315  -1.45  0.146
## UniqueAuthors2    0.52104    0.03443  15.13 < 2e-16 ***
## UniqueAuthors3    0.60603    0.04248  14.27 < 2e-16 ***
## UniqueAuthors4    0.64474    0.05506  11.71 < 2e-16 ***
## UniqueAuthors5    0.62725    0.06114  10.26 < 2e-16 ***
## Year1997          0.05182    0.13182   0.39  0.694
## Year1998          0.27824    0.13491   2.06  0.039 *
## Year1999          0.03806    0.12656   0.30  0.764
```

```

## Year2000      0.06303      0.12415      0.51      0.612
## Year2001      0.05881      0.13814      0.43      0.670
## Year2002     -0.06252      0.11970     -0.52      0.602
## Year2003      0.00703      0.11010      0.06      0.949
## Year2004     -0.09243      0.11072     -0.83      0.404
## Year2005     -0.13934      0.10952     -1.27      0.203
## Year2006     -0.00125      0.11443     -0.01      0.991
## Year2007     -0.08475      0.11246     -0.75      0.451
## Year2008     -0.05163      0.10759     -0.48      0.631
## Year2009     -0.09283      0.10625     -0.87      0.382
## Year2010      0.03788      0.10875      0.35      0.728
## Year2011     -0.05373      0.10585     -0.51      0.612
## Year2012     -0.01358      0.10484     -0.13      0.897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.173, Adjusted R-squared:  0.166
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 187 weights are ~= 1. The remaining 2417 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0767 0.8820 0.9390 0.9030 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.248 1      1.117
## LastAuthorFemale 1.235 1      1.111
## Year      1.048 16      1.001

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3803 4644346142 3.434 2004      2205      4      2.501
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1480 -0.5954 -0.0217  0.5381  2.5009
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.842275   0.091099   9.25  <2e-16 ***
## FirstAuthorFemale1 0.136420   0.050173   2.72   0.0066 **
## LastAuthorFemale1 -0.048440   0.053256  -0.91   0.3631
## Year1997         0.040001   0.130239   0.31   0.7588
## Year1998         0.305715   0.128931   2.37   0.0178 *
## Year1999         0.107433   0.124492   0.86   0.3882
## Year2000         0.198761   0.121763   1.63   0.1027
## Year2001         0.133240   0.139502   0.96   0.3396
## Year2002         0.010047   0.111951   0.09   0.9285
## Year2003         0.101832   0.106150   0.96   0.3375
## Year2004        -0.045598   0.108141  -0.42   0.6733
## Year2005        -0.130013   0.109846  -1.18   0.2367
```

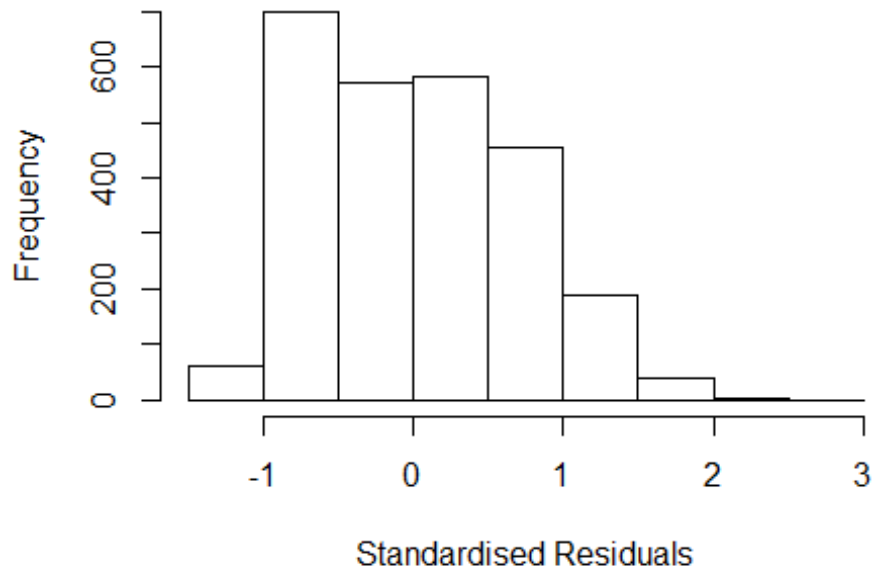
```

## Year2006          0.046451    0.112460    0.41    0.6796
## Year2007          -0.000748    0.108266   -0.01    0.9945
## Year2008          -0.026234    0.104826   -0.25    0.8024
## Year2009          -0.048087    0.102078   -0.47    0.6376
## Year2010           0.130918    0.103967    1.26    0.2081
## Year2011           0.057538    0.101714    0.57    0.5717
## Year2012           0.168538    0.099315    1.70    0.0898 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.784
## Multiple R-squared:  0.0197, Adjusted R-squared:  0.0129
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 2436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.287  0.895   0.945   0.925   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1      1.014
## Year              1.027 16      1.001

```



## Residuals from first author



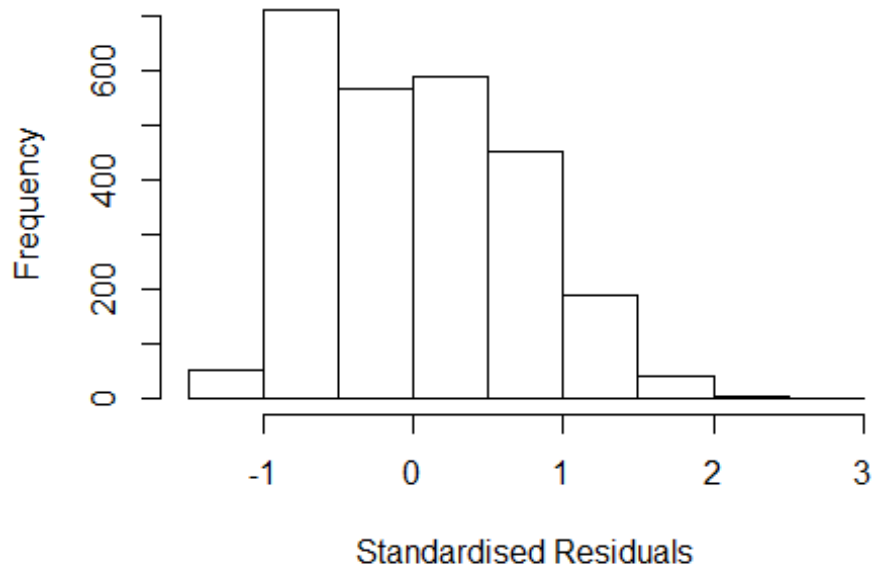
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3803 4644346142 3.434 2004      2205      4      2.501
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1521 -0.6062 -0.0202  0.5409  2.5253
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8.40e-01   9.15e-02   9.18  <2e-16 ***
## FirstAuthorFemale1 1.14e-01   4.62e-02   2.48   0.013 *
## Year1997        4.07e-02   1.31e-01   0.31   0.755
## Year1998        3.06e-01   1.29e-01   2.37   0.018 *
## Year1999        1.09e-01   1.25e-01   0.87   0.384
## Year2000        1.98e-01   1.22e-01   1.62   0.105
## Year2001        1.33e-01   1.40e-01   0.95   0.342
## Year2002         8.50e-03   1.12e-01   0.08   0.940
## Year2003        1.02e-01   1.07e-01   0.96   0.339
## Year2004       -4.53e-02   1.09e-01  -0.42   0.677
## Year2005       -1.30e-01   1.10e-01  -1.18   0.239
## Year2006        4.68e-02   1.13e-01   0.41   0.678
```

```

## Year2007          -8.81e-05    1.09e-01    0.00    0.999
## Year2008          -2.66e-02    1.05e-01   -0.25    0.800
## Year2009          -4.92e-02    1.02e-01   -0.48    0.631
## Year2010           1.27e-01    1.04e-01    1.22    0.221
## Year2011           5.76e-02    1.02e-01    0.56    0.573
## Year2012           1.68e-01    9.96e-02    1.69    0.091 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.784
## Multiple R-squared:  0.0194, Adjusted R-squared:  0.0129
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 2436 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.278  0.895  0.945  0.925  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1          1.008
## Year              1.017 16          1.001

```

## Residuals from last author



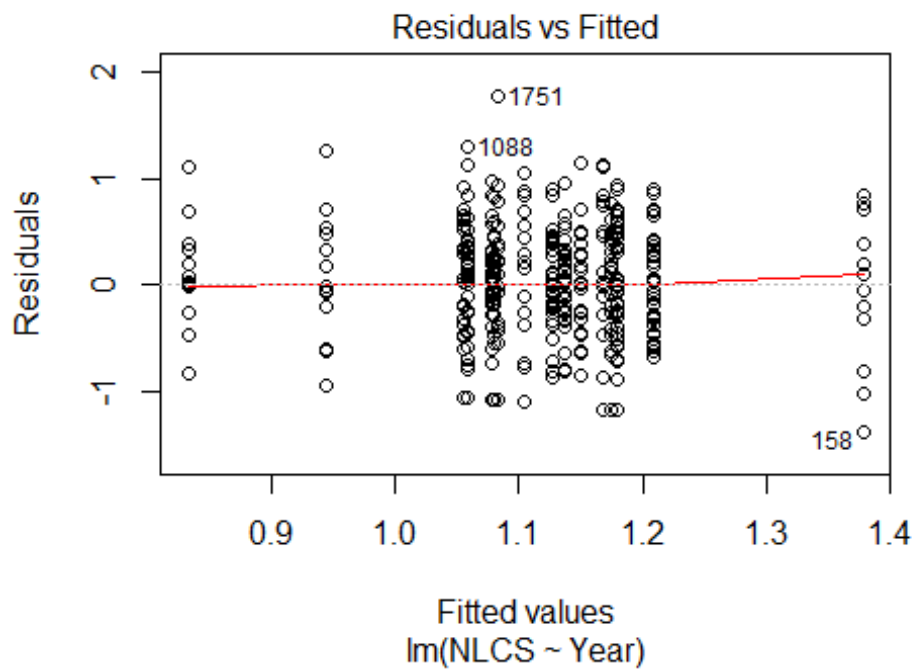
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3803 4644346142 3.434 2004      2205      4      2.501
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1689 -0.6046 -0.0199  0.5392  2.6276
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8424    0.0918   9.18  <2e-16 ***
## LastAuthorFemale1 0.0131    0.0474   0.28   0.783
## Year1997        0.0420    0.1304   0.32   0.747
## Year1998        0.3134    0.1293   2.42   0.015 *
## Year1999        0.1093    0.1247   0.88   0.381
## Year2000        0.1999    0.1225   1.63   0.103
## Year2001        0.1342    0.1400   0.96   0.338
## Year2002        0.0188    0.1124   0.17   0.867
## Year2003        0.1132    0.1069   1.06   0.290
## Year2004       -0.0360    0.1089  -0.33   0.741
## Year2005       -0.1166    0.1104  -1.06   0.291
## Year2006        0.0542    0.1130   0.48   0.631
```

```

## Year2007          0.0148      0.1088      0.14      0.892
## Year2008          -0.0206      0.1054     -0.19      0.845
## Year2009          -0.0449      0.1027     -0.44      0.662
## Year2010           0.1345      0.1047      1.28      0.199
## Year2011           0.0703      0.1023      0.69      0.492
## Year2012           0.1830      0.0999      1.83      0.067 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.785
## Multiple R-squared:  0.0171, Adjusted R-squared:  0.0106
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 169 weights are ~= 1. The remaining 2435 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.240  0.894  0.945  0.925  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2604"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2206"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 134  84 106  80  85 107 150 100 103 138 122 109 104  53  55
## 2011 2012
##  70  72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  18  13  19  14  14  19  36  22  22  36  42  23  37  20  22
## 2011 2012

```

```
## 25 15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 11 17 12 10 16 33 17 18 26 34 17 31 19 19
## 2011 2012
## 23 11
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```



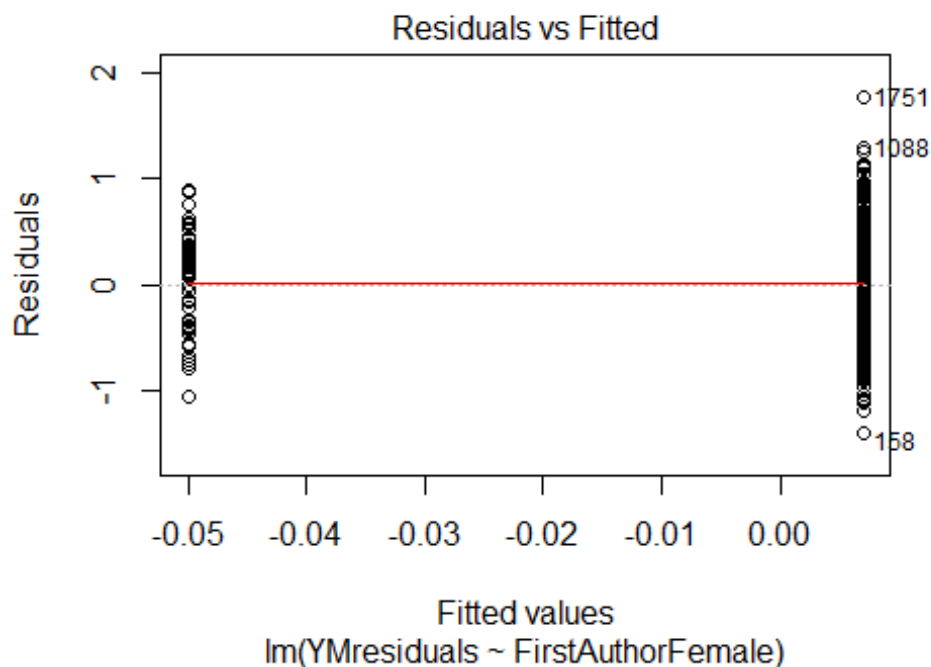
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 2.82842712474619"
## [1] "Male first author team size 2018 geometric mean: 2.36475008717522"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.82842712474619"
## [1] "Male last author team size 2018 geometric mean: 2.36475008717522"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

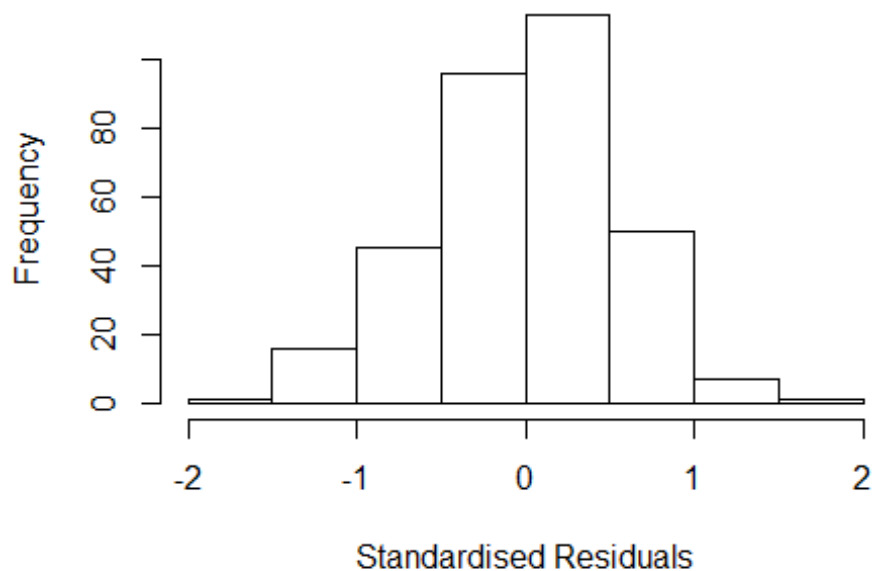
## Warning in lmrob.S(x, y, control = control, mf = mf): S refinements did
not
## converge (to refine.tol=1e-07) in 1000 (= k.max) steps
```



```
## [1] "Regression 2: First author gender, Last author gender, Year as factors"
```

```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.152  1      1.073
## LastAuthorFemale  1.064  1      1.031
## Year              1.222 16      1.006
```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
```

```
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
```

```
##
```

```
## Call:
```

```
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data = AllScopusDataOlderFirstLastGendered,
```

```
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
```

```
## \--> method = "MM"
```

```
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -1.5039 -0.3682  0.0184  0.3623  1.7955
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06596    0.26958   3.95 9.5e-05 ***
## FirstAuthorFemale1  0.05358    0.08520   0.63  0.530
## LastAuthorFemale1 -0.22983    0.12823  -1.79  0.074 .
## Year1997         0.43794    0.39468   1.11  0.268
## Year1998         0.16211    0.29399   0.55  0.582
```

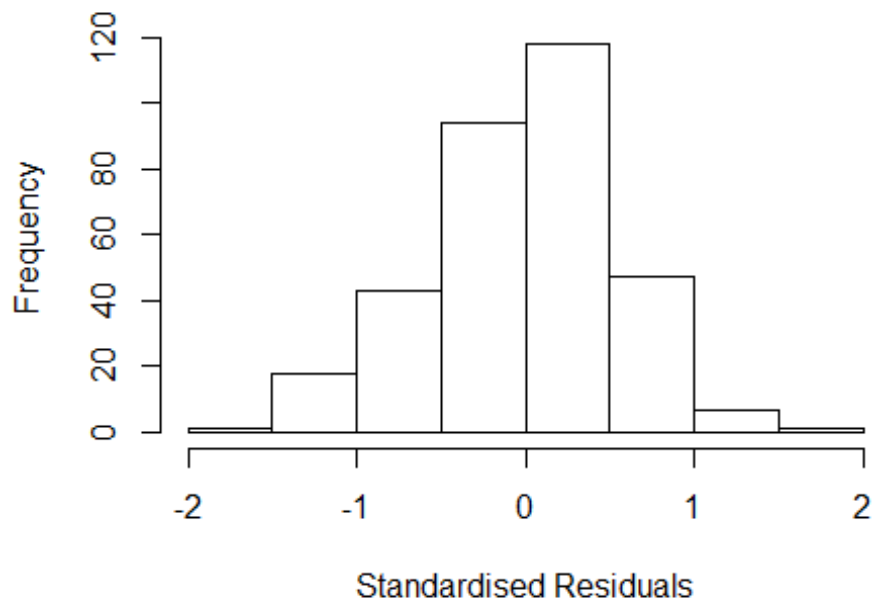
```

## Year1999      0.05444      0.31391      0.17      0.862
## Year2000     -0.14270      0.34703     -0.41      0.681
## Year2001      0.06893      0.35238      0.20      0.845
## Year2002      0.16347      0.28128      0.58      0.562
## Year2003      0.09596      0.30110      0.32      0.750
## Year2004     -0.06303      0.30200     -0.21      0.835
## Year2005      0.11040      0.28645      0.39      0.700
## Year2006      0.16071      0.28907      0.56      0.579
## Year2007      0.03191      0.29786      0.11      0.915
## Year2008      0.07254      0.28051      0.26      0.796
## Year2009     -0.01144      0.32316     -0.04      0.972
## Year2010     -0.01760      0.29997     -0.06      0.953
## Year2011      0.00539      0.28596      0.02      0.985
## Year2012     -0.25650      0.29666     -0.86      0.388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.0551, Adjusted R-squared:  0.000231
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 296 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.295  0.880  0.954  0.912  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.143  1      1.069
## Year              1.143 16      1.004

```



## Residuals from first author



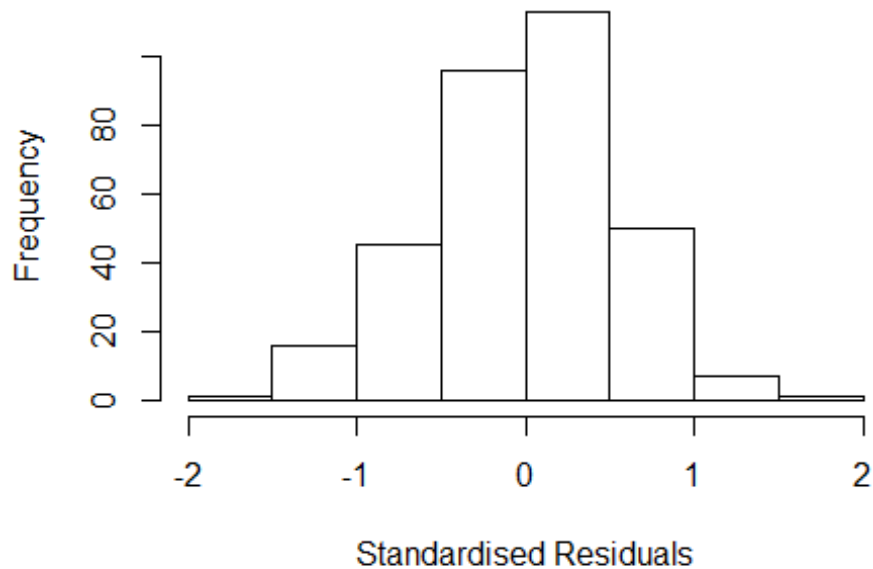
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5051 -0.3693 0.0267 0.3622 1.8087
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0658 0.2704 3.94 0.0001 ***
## FirstAuthorFemale1 0.0205 0.0874 0.23 0.8149
## Year1997 0.4393 0.3957 1.11 0.2678
## Year1998 0.1551 0.2970 0.52 0.6019
## Year1999 0.0546 0.3147 0.17 0.8624
## Year2000 -0.1593 0.3582 -0.44 0.6568
## Year2001 0.0694 0.3588 0.19 0.8468
## Year2002 0.1570 0.2826 0.56 0.5789
## Year2003 0.0815 0.3010 0.27 0.7867
## Year2004 -0.0789 0.3028 -0.26 0.7945
## Year2005 0.0962 0.2859 0.34 0.7368
## Year2006 0.1588 0.2894 0.55 0.5835
```

```

## Year2007          0.0193      0.3000      0.06      0.9487
## Year2008          0.0554      0.2805      0.20      0.8435
## Year2009         -0.0245      0.3215     -0.08      0.9393
## Year2010         -0.0420      0.3017     -0.14      0.8895
## Year2011         -0.0145      0.2867     -0.05      0.9597
## Year2012         -0.2661      0.3019     -0.88      0.3788
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0453, Adjusted R-squared:  -0.00684
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 33 weights are ~= 1. The remaining 296 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.873   0.952   0.910   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.068 1          1.033
## Year            1.068 16          1.002

```

## Residuals from last author



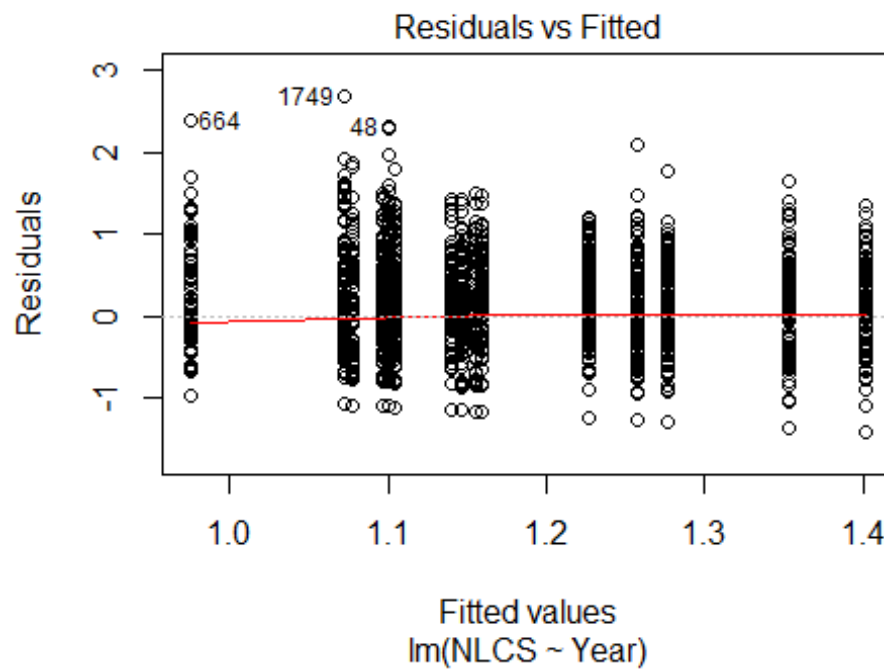
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5069 -0.3625 0.0114 0.3522 1.7891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06549 0.27134 3.93 0.00011 ***
## LastAuthorFemale1 -0.21687 0.12950 -1.67 0.09500 .
## Year1997 0.44141 0.39543 1.12 0.26517
## Year1998 0.17312 0.29465 0.59 0.55727
## Year1999 0.05480 0.31550 0.17 0.86223
## Year2000 -0.14396 0.34960 -0.41 0.68077
## Year2001 0.07701 0.35298 0.22 0.82745
## Year2002 0.16334 0.28315 0.58 0.56445
## Year2003 0.09912 0.30305 0.33 0.74384
## Year2004 -0.05134 0.30279 -0.17 0.86547
## Year2005 0.11663 0.28776 0.41 0.68555
## Year2006 0.16796 0.29016 0.58 0.56312
```

```

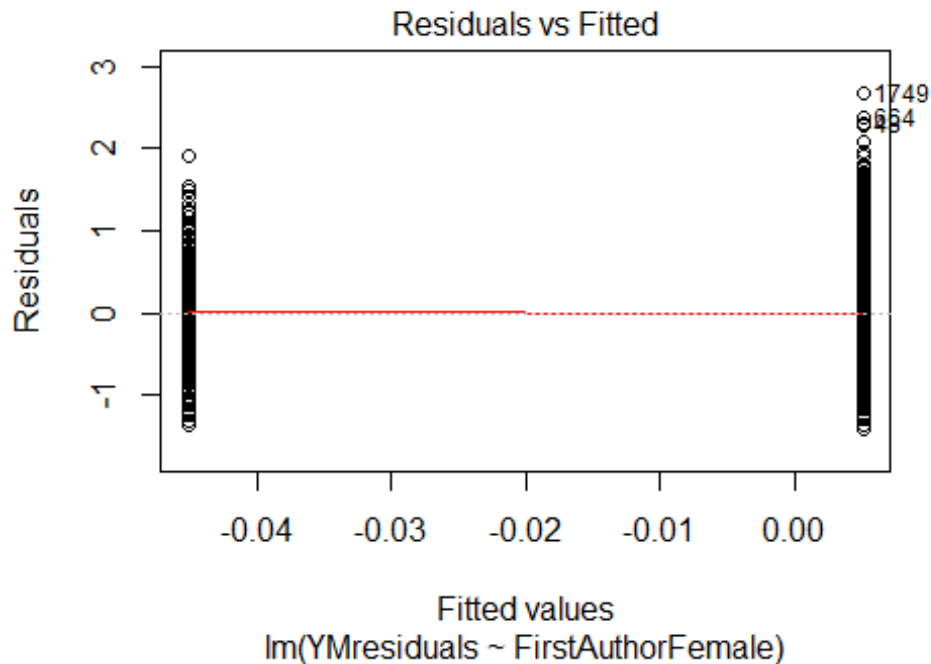
## Year2007          0.04421      0.29837      0.15  0.88231
## Year2008          0.08270      0.28138      0.29  0.76902
## Year2009         -0.00461      0.32411     -0.01  0.98867
## Year2010         -0.00423      0.29969     -0.01  0.98875
## Year2011          0.00871      0.28792      0.03  0.97588
## Year2012         -0.24869      0.29763     -0.84  0.40403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.0549, Adjusted R-squared:  0.00322
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 297 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.289  0.875   0.955   0.910   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.04e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 329"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2207"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   306  226  320  299  348  313  267  245  223  279  296  392  323  330  335
## 2011 2012
##   294  318
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    89   68   96   91  100   84   88   91   83  112  125  158  140  150  170
## 2011 2012

```

```
## 141 163
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 65 91 81 92 79 78 83 70 96 103 132 117 127 144
## 2011 2012
## 121 140
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.002
```

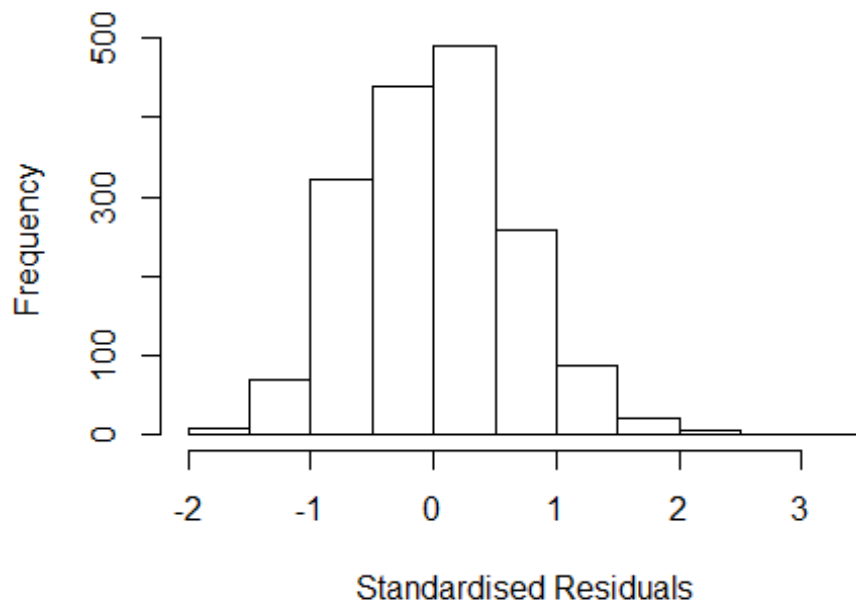


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.14, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 2.85458852909485"
## [1] "Male first author team size 2018 geometric mean: 2.58177964892625"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.14112736833832"
## [1] "Male last author team size 2018 geometric mean: 2.6916813533948"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.206 1      1.098
## LastAuthorFemale  1.219 1      1.104
## UniqueAuthors    1.251 4      1.028
## Year              1.347 16      1.009
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1749 0001224048 3.749 2001      1702      4      3.004
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5523 -0.4572  0.0151  0.4460  3.0043
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.79831    0.09819   8.13 8.2e-16 ***
## FirstAuthorFemale1 -0.05305    0.06166  -0.86  0.390
## LastAuthorFemale1 -0.02348    0.05997  -0.39  0.695
## UniqueAuthors2     0.47395    0.04294  11.04 < 2e-16 ***
## UniqueAuthors3     0.47939    0.04710  10.18 < 2e-16 ***
## UniqueAuthors4     0.69732    0.06193  11.26 < 2e-16 ***
## UniqueAuthors5     0.52051    0.07298   7.13 1.5e-12 ***
## Year1997        -0.00620    0.14416  -0.04  0.966
## Year1998        -0.15891    0.12602  -1.26  0.207
## Year1999        -0.02862    0.12339  -0.23  0.817
```

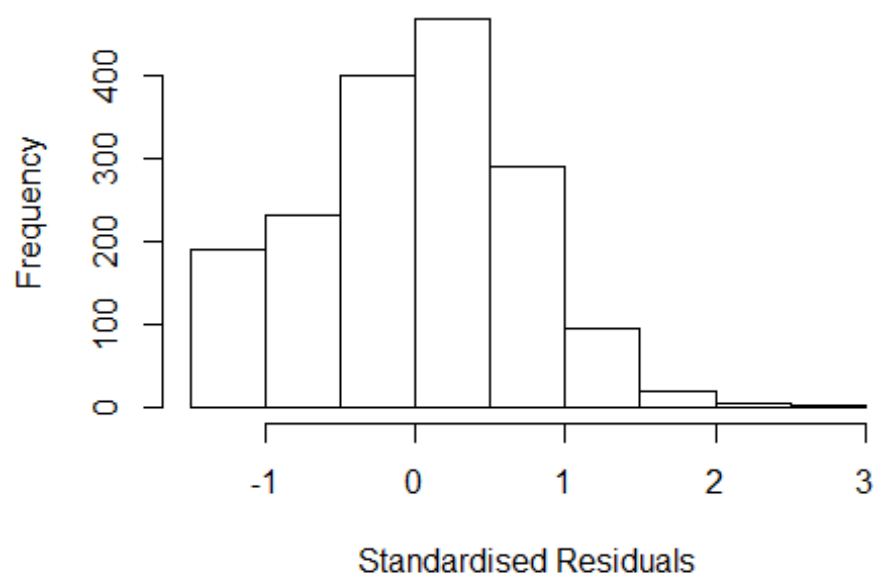
```

## Year2000      -0.08326    0.12269   -0.68    0.497
## Year2001      -0.05365    0.12248   -0.44    0.661
## Year2002      -0.05908    0.12953   -0.46    0.648
## Year2003       0.01289    0.11931    0.11    0.914
## Year2004      -0.04149    0.13170   -0.32    0.753
## Year2005       0.26058    0.11711    2.23    0.026 *
## Year2006       0.27457    0.11671    2.35    0.019 *
## Year2007      -0.01123    0.11396   -0.10    0.921
## Year2008       0.08146    0.11431    0.71    0.476
## Year2009       0.00218    0.11327    0.02    0.985
## Year2010      -0.11674    0.11050   -1.06    0.291
## Year2011       0.01099    0.11509    0.10    0.924
## Year2012       0.03190    0.10829    0.29    0.768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.649
## Multiple R-squared:  0.155, Adjusted R-squared:  0.144
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8740 0.9480 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.188 1      1.090
## LastAuthorFemale  1.180 1      1.086
## Year              1.073 16      1.002

```



## Residuals from first and last author

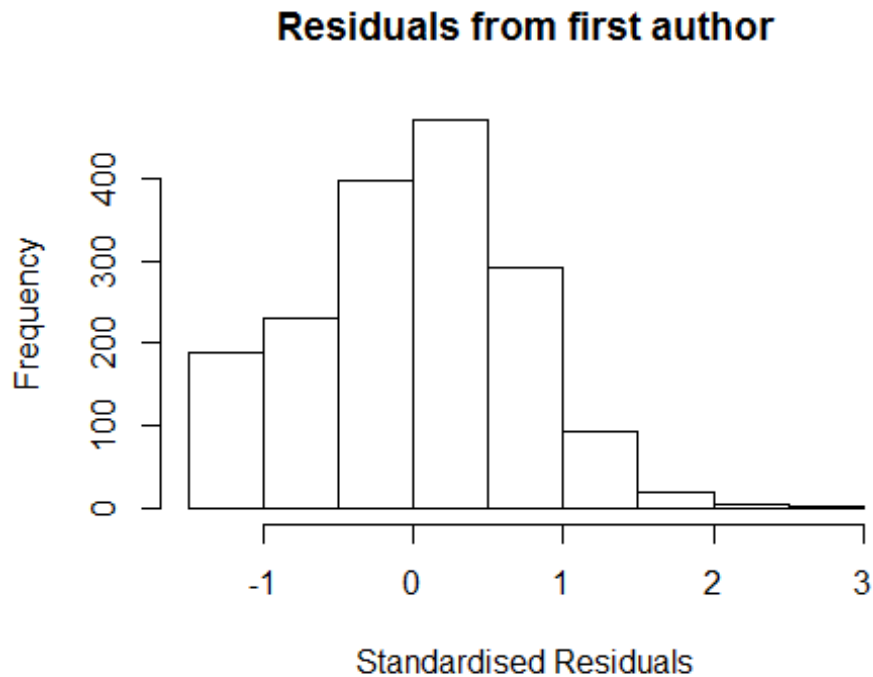


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1749 0001224048 3.749 2001      1702      4      2.729
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4330 -0.4944  0.0357  0.4783  2.7293
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0774     0.1066   10.11  <2e-16 ***
## FirstAuthorFemale1 -0.0501     0.0662   -0.76   0.4496
## LastAuthorFemale1 -0.0318     0.0638   -0.50   0.6180
## Year1997          -0.0617     0.1587   -0.39   0.6972
## Year1998          -0.1441     0.1350   -1.07   0.2857
## Year1999           0.0376     0.1331    0.28   0.7774
## Year2000          -0.0395     0.1315   -0.30   0.7640
## Year2001          -0.0577     0.1333   -0.43   0.6651
## Year2002          -0.0550     0.1404   -0.39   0.6954
## Year2003           0.0515     0.1298    0.40   0.6917
## Year2004           0.0140     0.1411    0.10   0.9207
## Year2005           0.3393     0.1268    2.67   0.0075 **
```

```

## Year2006          0.3556      0.1279      2.78      0.0055 **
## Year2007          0.0350      0.1263      0.28      0.7820
## Year2008          0.1450      0.1210      1.20      0.2308
## Year2009          0.0781      0.1233      0.63      0.5265
## Year2010         -0.0221      0.1194     -0.19      0.8529
## Year2011          0.1389      0.1248      1.11      0.2656
## Year2012          0.1947      0.1171      1.66      0.0965 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.035, Adjusted R-squared:  0.0247
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1553 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.107  0.867   0.950   0.913   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.04 1          1.020
## Year              1.04 16          1.001

```

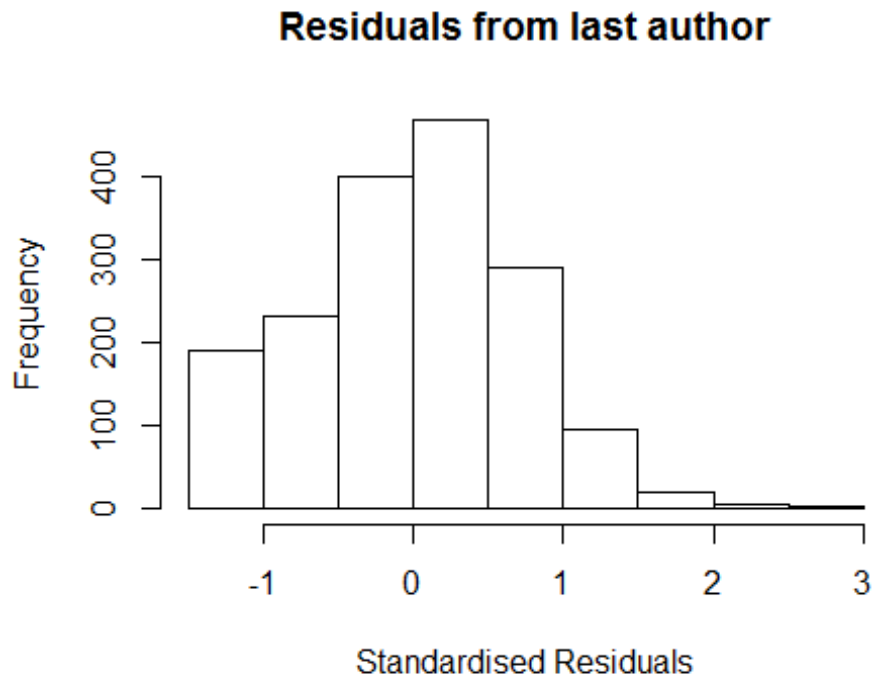


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1749 0001224048 3.749 2001      1702      4      2.729
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4297 -0.4936  0.0371  0.4811  2.7305
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0766     0.1063   10.13  <2e-16 ***
## FirstAuthorFemale1 -0.0617     0.0625  -0.99   0.3233
## Year1997         -0.0638     0.1582  -0.40   0.6869
## Year1998         -0.1460     0.1345  -1.09   0.2780
## Year1999          0.0360     0.1329   0.27   0.7864
## Year2000         -0.0402     0.1313  -0.31   0.7595
## Year2001         -0.0582     0.1331  -0.44   0.6620
## Year2002         -0.0547     0.1403  -0.39   0.6965
## Year2003          0.0513     0.1296   0.40   0.6924
## Year2004          0.0125     0.1410   0.09   0.9291
## Year2005          0.3374     0.1265   2.67   0.0077 **
## Year2006          0.3530     0.1274   2.77   0.0056 **
```

```

## Year2007          0.0325      0.1258      0.26      0.7963
## Year2008          0.1448      0.1207      1.20      0.2307
## Year2009          0.0768      0.1231      0.62      0.5329
## Year2010         -0.0234      0.1191     -0.20      0.8443
## Year2011          0.1382      0.1246      1.11      0.2673
## Year2012          0.1942      0.1168      1.66      0.0964 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.71
## Multiple R-squared:  0.0349, Adjusted R-squared:  0.0251
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1553 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.867   0.949   0.913   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1      1.016
## Year              1.032 16      1.001

```



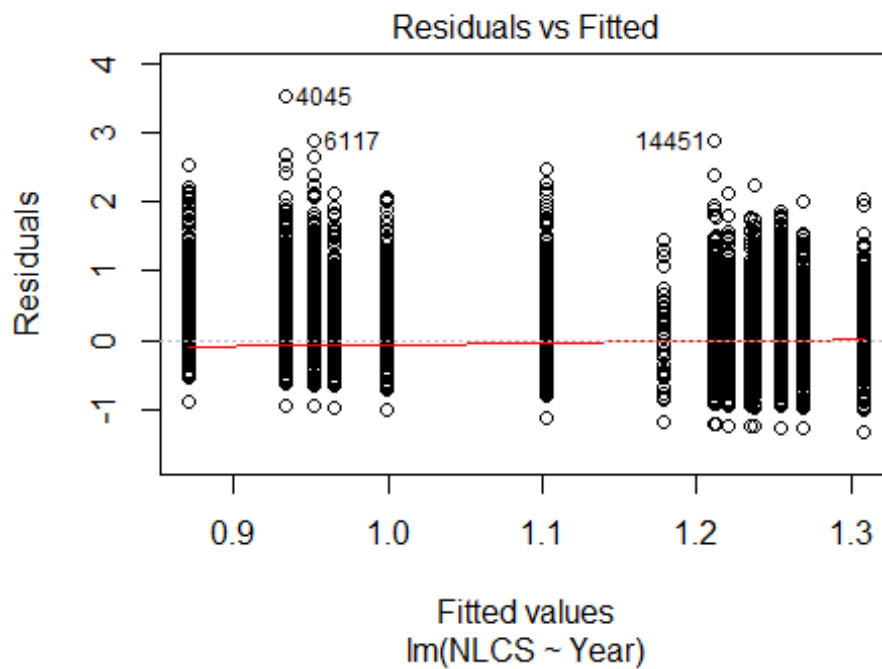
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1749 0001224048 3.749 2001      1702      4      2.729
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4300 -0.4903  0.0382  0.4770  2.7341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0733     0.1063  10.10  <2e-16 ***
## LastAuthorFemale1 -0.0506     0.0606  -0.83   0.4041
## Year1997         -0.0608     0.1583  -0.38   0.7009
## Year1998         -0.1409     0.1348  -1.05   0.2959
## Year1999          0.0408     0.1328   0.31   0.7588
## Year2000         -0.0388     0.1315  -0.29   0.7681
## Year2001         -0.0584     0.1335  -0.44   0.6618
## Year2002         -0.0517     0.1403  -0.37   0.7127
## Year2003          0.0498     0.1297   0.38   0.7013
## Year2004          0.0153     0.1412   0.11   0.9137
## Year2005          0.3382     0.1269   2.66   0.0078 **
## Year2006          0.3567     0.1282   2.78   0.0055 **
```

```

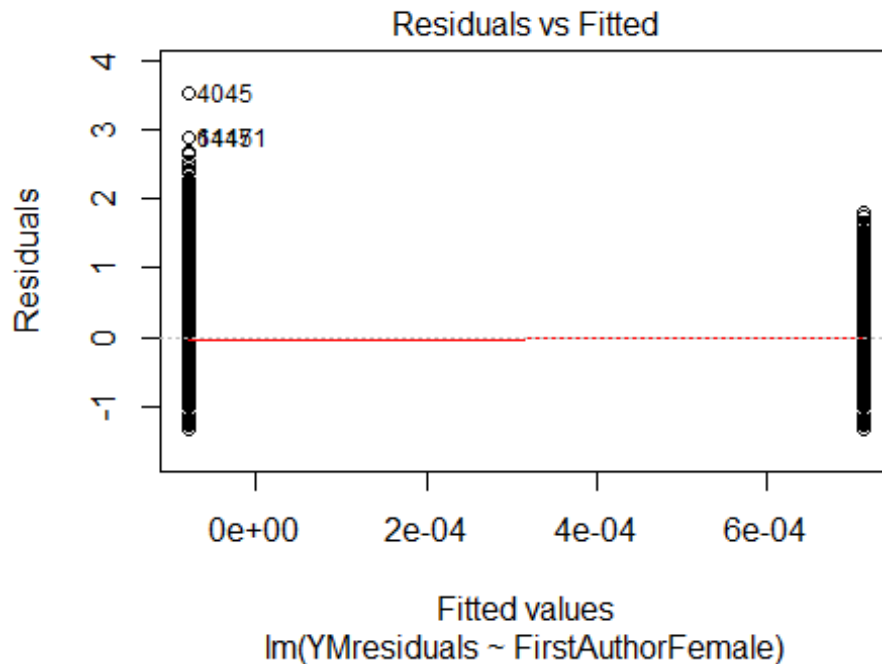
## Year2007          0.0356      0.1263      0.28      0.7783
## Year2008          0.1468      0.1211      1.21      0.2256
## Year2009          0.0779      0.1233      0.63      0.5273
## Year2010         -0.0207      0.1193     -0.17      0.8625
## Year2011          0.1393      0.1247      1.12      0.2639
## Year2012          0.1965      0.1170      1.68      0.0933 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.709
## Multiple R-squared:  0.0346, Adjusted R-squared:  0.0249
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1561 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.104  0.866  0.950  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.89e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1697"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2208"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2167 2033 1896 213 1590 1337 1101 947 883 970 1158 1205 1183 1231 1193
## 2011 2012
## 1166 1170
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 464 449 467 46 415 311 368 283 322 369 393 449 448 526 481
## 2011 2012

```

```
## 487 519
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 410 412 423 41 383 272 313 246 277 315 331 371 367 427 407
## 2011 2012
## 411 425
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 160, df = 16, p-value <2e-16
```



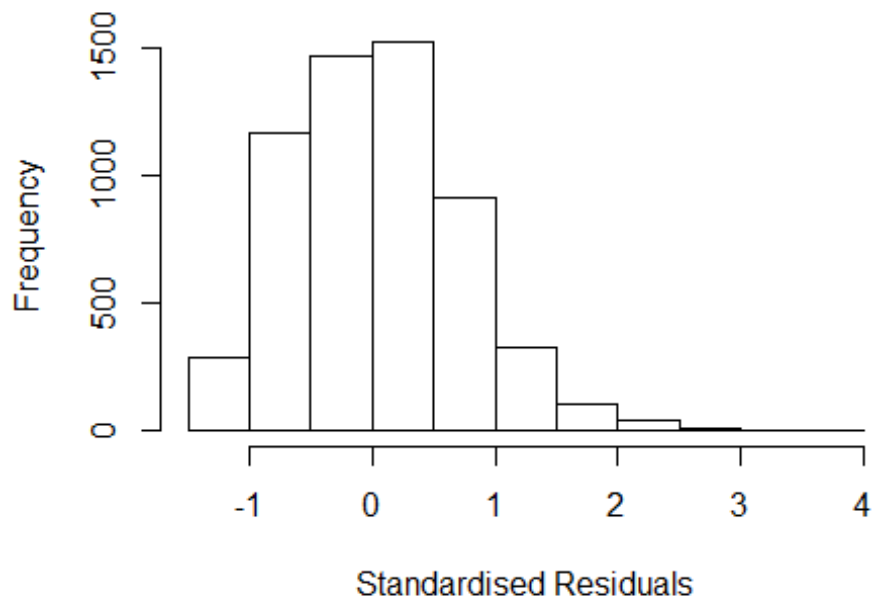
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 3.66488864891151"
## [1] "Male first author team size 2018 geometric mean: 3.13175370026021"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.32080140031289"
## [1] "Male last author team size 2018 geometric mean: 3.13887055487225"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061 1          1.030
## LastAuthorFemale  1.054 1          1.027
## UniqueAuthors    1.190 4          1.022
## Year             1.201 16          1.006
```



## Residuals from first and last author and team size



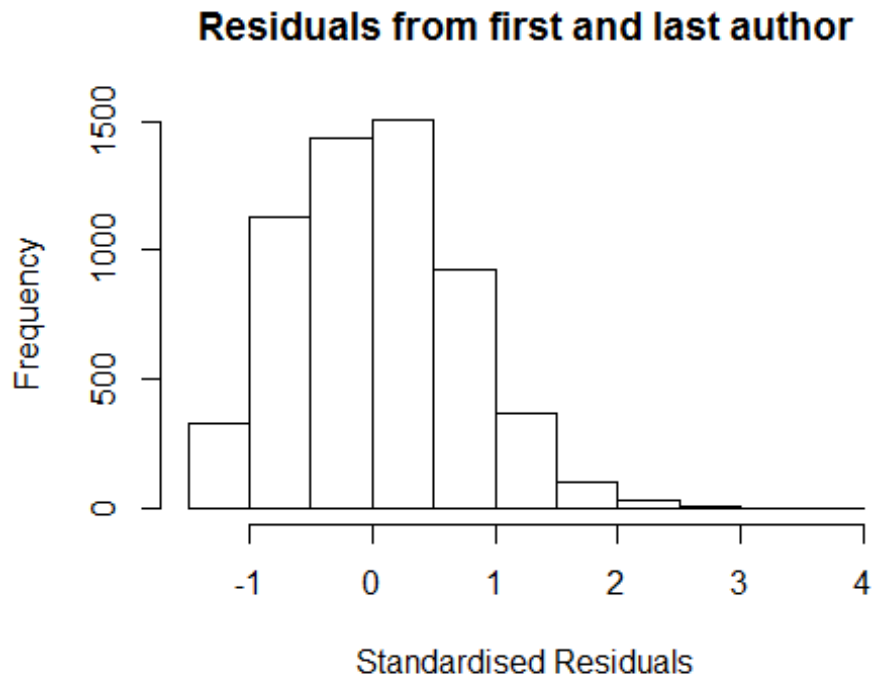
```
## [1] "List of 8 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 413   0030193785 3.070 1996    1900      2    2.526
## 2712  0031079076 3.600 1997    2102      2    2.603
## 2835  0031200253 3.451 1997    2102      2    2.822
## 4045  0346613481 4.450 1997    2208      1    3.821
## 5226  0032164263 3.590 1998    1900      2    2.925
## 6117  0032070303 3.831 1998    1705      2    2.798
## 8069  0034187380 3.580 2000    2208      2    2.779
## 14451 0742290132 4.094 2004    2208      2    3.209
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.45870 -0.49652 -0.00181  0.46607  3.82084
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.5437      0.0427   12.75 < 2e-16 ***
## FirstAuthorFemale1 -0.0109      0.0289   -0.38  0.705
## LastAuthorFemale1  -0.0759      0.0395   -1.92  0.055 .
##
```

```

## UniqueAuthors2      0.3682      0.0271     13.59 < 2e-16 ***
## UniqueAuthors3      0.3927      0.0285     13.79 < 2e-16 ***
## UniqueAuthors4      0.4470      0.0335     13.33 < 2e-16 ***
## UniqueAuthors5      0.4824      0.0321     15.04 < 2e-16 ***
## Year1997             0.0855      0.0576      1.49    0.138
## Year1998             0.1214      0.0580      2.09    0.036 *
## Year1999             0.4075      0.1239      3.29    0.001 **
## Year2000             0.2571      0.0590      4.36    1.4e-05 ***
## Year2001             0.0602      0.0622      0.97    0.333
## Year2002             0.1286      0.0580      2.22    0.027 *
## Year2003             0.4236      0.0637      6.66    3.1e-11 ***
## Year2004             0.3413      0.0593      5.75    9.3e-09 ***
## Year2005             0.4681      0.0565      8.29    < 2e-16 ***
## Year2006             0.3966      0.0567      6.99    3.0e-12 ***
## Year2007             0.3467      0.0541      6.41    1.6e-10 ***
## Year2008             0.3640      0.0550      6.61    4.1e-11 ***
## Year2009             0.3383      0.0529      6.39    1.8e-10 ***
## Year2010             0.3297      0.0527      6.26    4.1e-10 ***
## Year2011             0.3585      0.0530      6.77    1.4e-11 ***
## Year2012             0.3349      0.0525      6.37    2.0e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.119, Adjusted R-squared:  0.115
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(720,2729) are outliers with |weight| = 0 ( < 1.7e-05);
## 476 weights are ~= 1. The remaining 5353 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0275 0.8690 0.9450 0.9070 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

```

```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1      1.015
## LastAuthorFemale  1.035 1      1.017
## Year              1.021 16      1.001
```



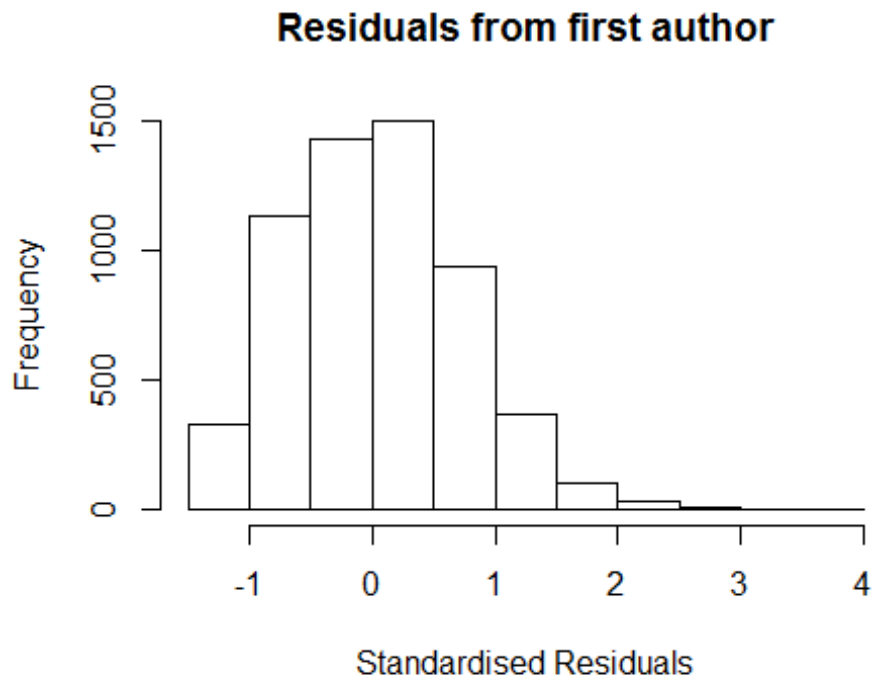
```
## [1] "List of 8 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 882   0030196132 3.406 1996    2207      3      2.657
## 2712  0031079076 3.600 1997    2102      2      2.754
## 2835  0031200253 3.451 1997    2102      2      2.605
## 4045  0346613481 4.450 1997    2208      1      3.604
## 5226  0032164263 3.590 1998    1900      2      2.717
## 6117  0032070303 3.831 1998    1705      2      2.958
## 8069  0034187380 3.580 2000    2208      2      2.529
## 14451 0742290132 4.094 2004    2208      2      2.901
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34654 -0.49961  0.00718  0.49360  3.60428
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept)          0.7495      0.0443    16.93 < 2e-16 ***
## FirstAuthorFemale1   0.0301      0.0292     1.03  0.3018
## LastAuthorFemale1    -0.0829     0.0399    -2.08  0.0379 *
## Year1997             0.0963      0.0605     1.59  0.1116
## Year1998             0.1239      0.0615     2.01  0.0441 *
## Year1999             0.4475      0.1150     3.89  0.0001 ***
## Year2000             0.3012      0.0607     4.96  7.3e-07 ***
## Year2001             0.1601      0.0650     2.46  0.0138 *
## Year2002             0.2248      0.0603     3.73  0.0002 ***
## Year2003             0.5191      0.0668     7.77  9.4e-15 ***
## Year2004             0.4435      0.0616     7.20  6.6e-13 ***
## Year2005             0.5669      0.0585     9.69 < 2e-16 ***
## Year2006             0.5011      0.0583     8.60 < 2e-16 ***
## Year2007             0.4695      0.0574     8.18  3.4e-16 ***
## Year2008             0.4789      0.0570     8.40 < 2e-16 ***
## Year2009             0.4773      0.0544     8.78 < 2e-16 ***
## Year2010             0.4573      0.0545     8.39 < 2e-16 ***
## Year2011             0.4933      0.0548     9.00 < 2e-16 ***
## Year2012             0.4912      0.0539     9.11 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.713
## Multiple R-squared:  0.0596, Adjusted R-squared:  0.0567
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 720 is an outlier with |weight| = 0 ( < 1.7e-05);
## 451 weights are ~ = 1. The remaining 5379 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0465 0.8680 0.9490 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))

```

```
## FirstAuthorFemale 1.008 1 1.004
## Year 1.008 16 1.000
```



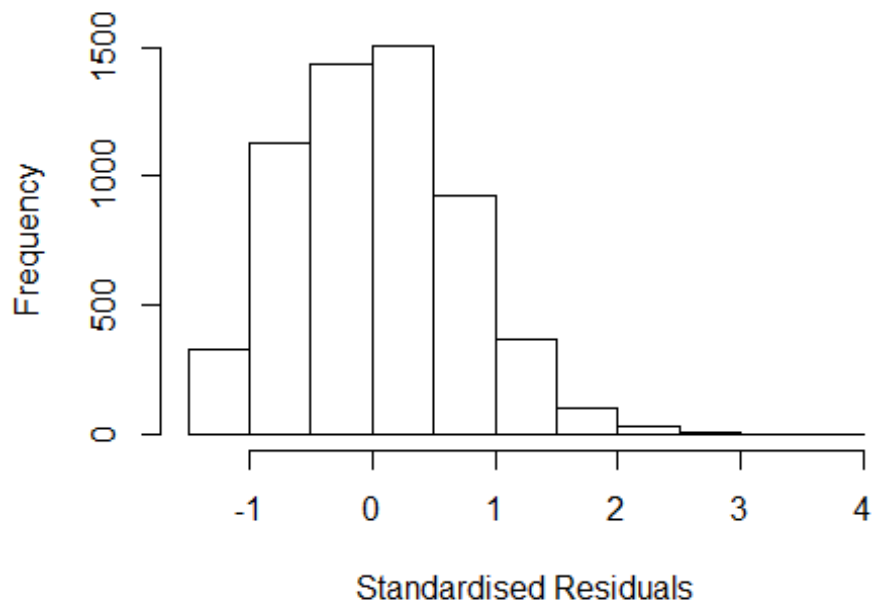
```
## [1] "List of 8 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 882   0030196132 3.406 1996    2207      3      2.657
## 2712  0031079076 3.600 1997    2102      2      2.754
## 2835  0031200253 3.451 1997    2102      2      2.605
## 4045  0346613481 4.450 1997    2208      1      3.604
## 5226  0032164263 3.590 1998    1900      2      2.717
## 6117  0032070303 3.831 1998    1705      2      2.958
## 8069  0034187380 3.580 2000    2208      2      2.529
## 14451 0742290132 4.094 2004    2208      2      2.901
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32622 -0.50117  0.00959  0.49610  3.60903
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7479    0.0443   16.90  < 2e-16 ***
## FirstAuthorFemale1 0.0134    0.0295    0.46  0.64907
```

```

## Year1997          0.0931      0.0604      1.54  0.12336
## Year1998          0.1187      0.0615      1.93  0.05358 .
## Year1999          0.4429      0.1162      3.81  0.00014 ***
## Year2000          0.3005      0.0607      4.95  7.7e-07 ***
## Year2001          0.1585      0.0651      2.44  0.01491 *
## Year2002          0.2213      0.0602      3.67  0.00024 ***
## Year2003          0.5159      0.0668      7.72  1.3e-14 ***
## Year2004          0.4404      0.0615      7.17  8.7e-13 ***
## Year2005          0.5649      0.0584      9.67  < 2e-16 ***
## Year2006          0.4990      0.0582      8.58  < 2e-16 ***
## Year2007          0.4659      0.0572      8.15  4.3e-16 ***
## Year2008          0.4764      0.0571      8.35  < 2e-16 ***
## Year2009          0.4761      0.0544      8.76  < 2e-16 ***
## Year2010          0.4570      0.0545      8.38  < 2e-16 ***
## Year2011          0.4900      0.0547      8.96  < 2e-16 ***
## Year2012          0.4885      0.0538      9.08  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.714
## Multiple R-squared:  0.0587, Adjusted R-squared:  0.056
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 720 is an outlier with |weight| = 0 ( < 1.7e-05);
## 448 weights are ~= 1. The remaining 5382 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0463 0.8700 0.9490 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year              1.012 16          1.000

```

## Residuals from last author



```
## [1] "List of 8 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 882   0030196132 3.406 1996    2207      3    2.657
## 2712  0031079076 3.600 1997    2102      2    2.754
## 2835  0031200253 3.451 1997    2102      2    2.605
## 4045  0346613481 4.450 1997    2208      1    3.604
## 5226  0032164263 3.590 1998    1900      2    2.717
## 6117  0032070303 3.831 1998    1705      2    2.958
## 8069  0034187380 3.580 2000    2208      2    2.529
## 14451 0742290132 4.094 2004    2208      2    2.901
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.31907 -0.50142  0.00618  0.49266  3.60228
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7517    0.0442   17.01  < 2e-16 ***
## LastAuthorFemale1 -0.0740    0.0394   -1.88  0.06030 .
## Year1997         0.0960    0.0605    1.59  0.11264
## Year1998         0.1236    0.0615    2.01  0.04449 *
## Year1999         0.4449    0.1152    3.86  0.00011 ***
```

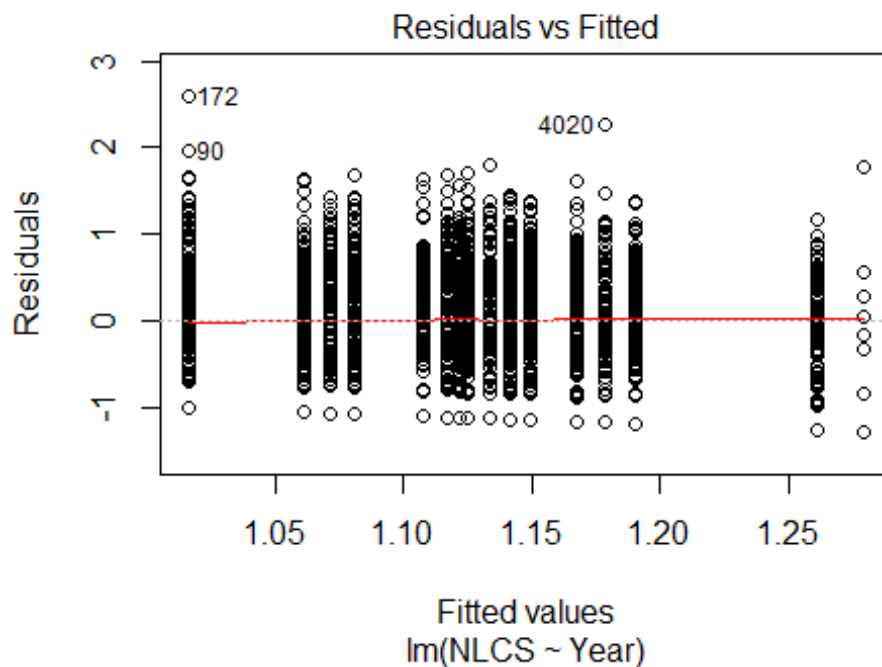
```

## Year2000          0.3013      0.0608      4.96  7.3e-07 ***
## Year2001          0.1597      0.0650      2.46  0.01402 *
## Year2002          0.2255      0.0604      3.74  0.00019 ***
## Year2003          0.5198      0.0668      7.78  8.7e-15 ***
## Year2004          0.4437      0.0616      7.21  6.5e-13 ***
## Year2005          0.5673      0.0585      9.69  < 2e-16 ***
## Year2006          0.5019      0.0583      8.61  < 2e-16 ***
## Year2007          0.4691      0.0574      8.18  3.6e-16 ***
## Year2008          0.4795      0.0570      8.41  < 2e-16 ***
## Year2009          0.4779      0.0544      8.79  < 2e-16 ***
## Year2010          0.4588      0.0545      8.41  < 2e-16 ***
## Year2011          0.4938      0.0548      9.01  < 2e-16 ***
## Year2012          0.4917      0.0540      9.11  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.713
## Multiple R-squared:  0.0595, Adjusted R-squared:  0.0567
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 720 is an outlier with |weight| = 0 ( < 1.7e-05);
## 446 weights are ~= 1. The remaining 5384 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0471 0.8680 0.9490 0.9100 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          1.71e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500          50          2          1          1000          200
## trace.lev mts compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5831"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2209"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##

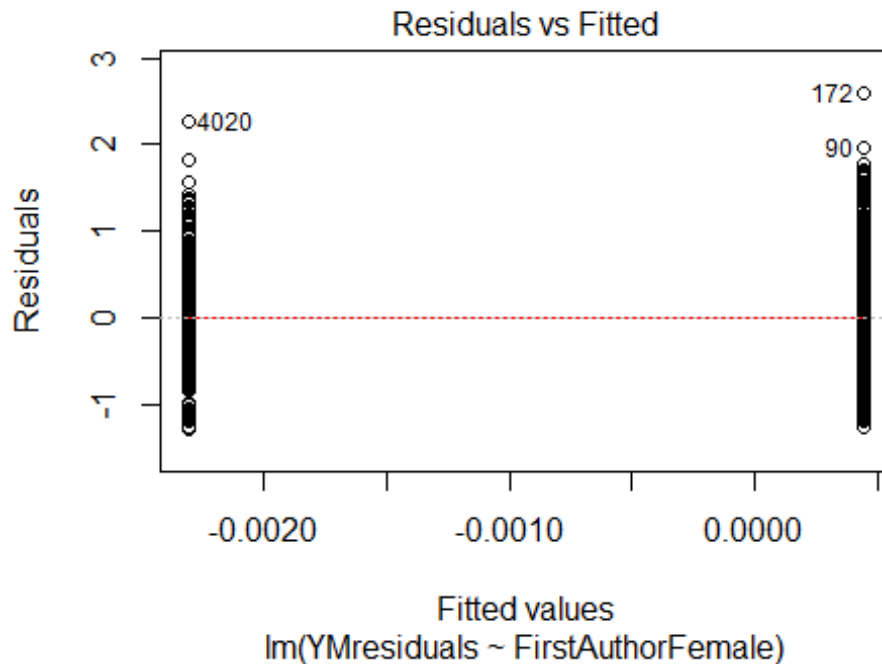
```



```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    38  475  480  484  565  507  370  356  353  326  433  442  401  424  449
## 2011 2012
##   407  418
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8  160  147  136  177  108  125  145  133  124  149  197  193  190  203
## 2011 2012
##   195  220
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8  147  134  114  154   94  113  131  112   92  123  170  156  170  171
## 2011 2012
##   159  185
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 42, df = 16, p-value = 4e-04
```

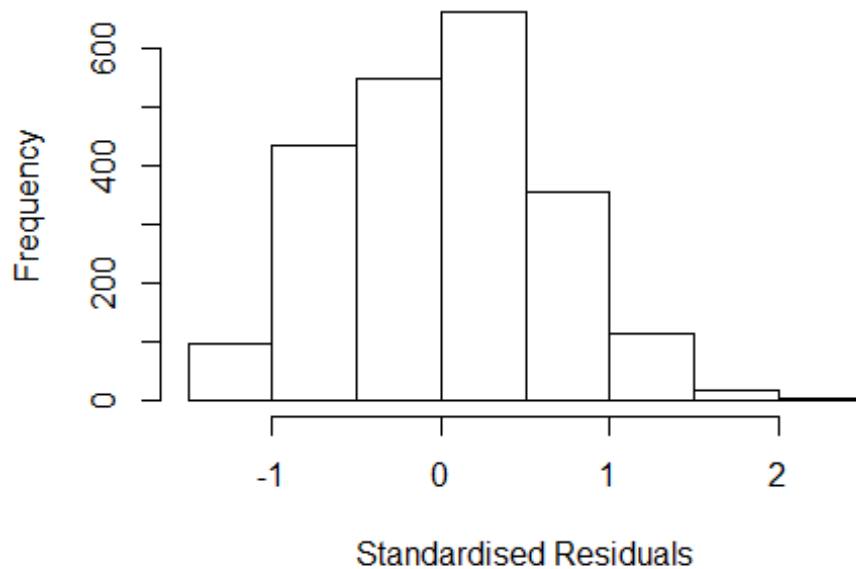


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.63154960345251"
## [1] "Male first author team size 2018 geometric mean: 3.02773511126461"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.54774927640468"
## [1] "Male last author team size 2018 geometric mean: 3.03509989107506"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.125 1 1.060
## LastAuthorFemale 1.114 1 1.056
## UniqueAuthors 1.207 4 1.024
## Year 1.241 16 1.007
```

## Residuals from first and last author and team size



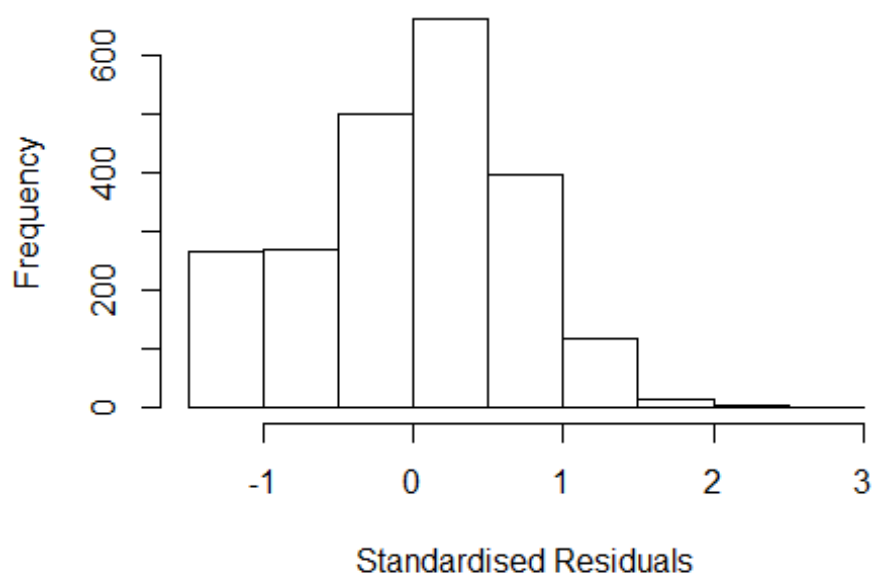
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.420 -0.466 0.024 0.439 2.412
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.89531 0.26222 3.41 0.00065 ***
## FirstAuthorFemale1 -0.01462 0.04157 -0.35 0.72512
## LastAuthorFemale1 0.01932 0.03959 0.49 0.62557
## UniqueAuthors2 0.41774 0.03963 10.54 < 2e-16 ***
## UniqueAuthors3 0.46203 0.03976 11.62 < 2e-16 ***
## UniqueAuthors4 0.51475 0.05061 10.17 < 2e-16 ***
## UniqueAuthors5 0.41950 0.05559 7.55 6.5e-14 ***
## Year1997 -0.11880 0.26903 -0.44 0.65883
## Year1998 0.03681 0.26947 0.14 0.89136
## Year1999 -0.11603 0.27210 -0.43 0.66984
```

```

## Year2000      -0.10031    0.26873   -0.37  0.70899
## Year2001      -0.06657    0.27277   -0.24  0.80720
## Year2002       0.00567    0.26956    0.02  0.98322
## Year2003      -0.03830    0.26792   -0.14  0.88635
## Year2004      -0.03907    0.26869   -0.15  0.88440
## Year2005      -0.01009    0.26977   -0.04  0.97018
## Year2006      -0.09355    0.27041   -0.35  0.72940
## Year2007      -0.16437    0.26667   -0.62  0.53770
## Year2008      -0.07653    0.26782   -0.29  0.77509
## Year2009      -0.12449    0.26861   -0.46  0.64308
## Year2010       0.00218    0.26643    0.01  0.99348
## Year2011      -0.08206    0.26750   -0.31  0.75906
## Year2012      -0.12628    0.26542   -0.48  0.63429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.656
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0933
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 218 weights are ~= 1. The remaining 2015 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.147  0.868  0.947   0.910   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.48e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.085 1      1.041
## LastAuthorFemale  1.079 1      1.039
## Year              1.062 16      1.002

```

## Residuals from first and last author



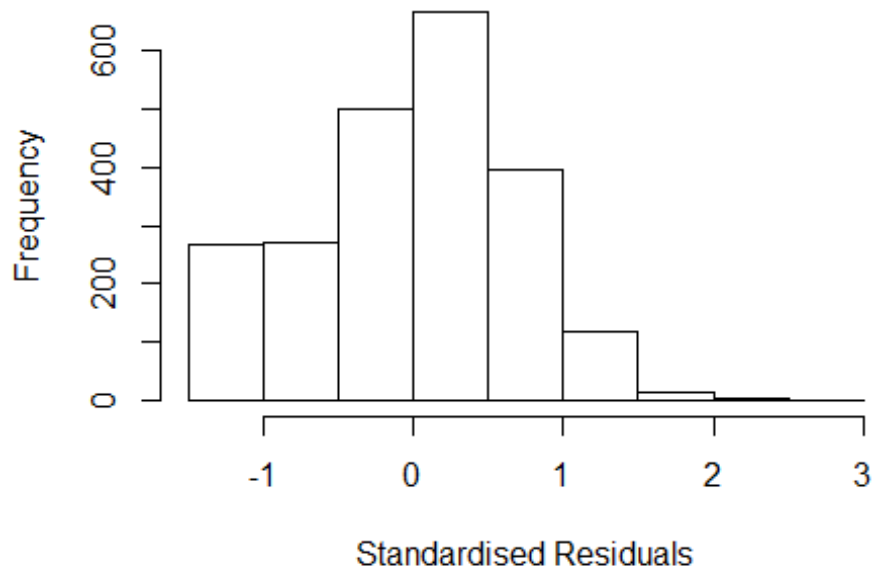
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 172 0031200344 3.606 1997    2102      2      2.629
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2509 -0.4722  0.0575  0.4708  2.6292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1875     0.3034   3.91 9.4e-05 ***
## FirstAuthorFemale1 -0.0169     0.0438  -0.38   0.70
## LastAuthorFemale1  0.0086     0.0419   0.21   0.84
## Year1997          -0.2107     0.3113  -0.68   0.50
## Year1998          -0.0515     0.3120  -0.16   0.87
## Year1999          -0.1389     0.3140  -0.44   0.66
## Year2000          -0.0853     0.3100  -0.28   0.78
## Year2001          -0.0776     0.3135  -0.25   0.80
## Year2002          -0.0132     0.3115  -0.04   0.97
## Year2003          -0.0473     0.3093  -0.15   0.88
## Year2004          -0.0138     0.3101  -0.04   0.96
## Year2005           0.0548     0.3105   0.18   0.86
```

```

## Year2006          -0.0927      0.3120    -0.30      0.77
## Year2007          -0.1454      0.3083    -0.47      0.64
## Year2008          -0.0460      0.3089    -0.15      0.88
## Year2009          -0.1177      0.3097    -0.38      0.70
## Year2010           0.0197      0.3078     0.06      0.95
## Year2011          -0.0475      0.3084    -0.15      0.88
## Year2012          -0.0520      0.3063    -0.17      0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.683
## Multiple R-squared:  0.00844,    Adjusted R-squared:  0.000378
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 189 weights are ~= 1. The remaining 2044 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.106  0.863  0.947   0.911   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.48e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.04 1      1.020
## Year              1.04 16      1.001

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 172 0031200344 3.606 1997    2102      2      2.629
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2430 -0.4731  0.0577  0.4699  2.6285
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1872     0.3037   3.91 9.6e-05 ***
## FirstAuthorFemale1 -0.0137     0.0433  -0.32  0.75
## Year1997          -0.2097     0.3117  -0.67  0.50
## Year1998          -0.0505     0.3124  -0.16  0.87
## Year1999          -0.1382     0.3143  -0.44  0.66
## Year2000          -0.0847     0.3104  -0.27  0.78
## Year2001          -0.0768     0.3137  -0.24  0.81
## Year2002          -0.0121     0.3118  -0.04  0.97
## Year2003          -0.0461     0.3097  -0.15  0.88
## Year2004          -0.0122     0.3104  -0.04  0.97
## Year2005           0.0558     0.3108   0.18  0.86
## Year2006          -0.0912     0.3123  -0.29  0.77
```

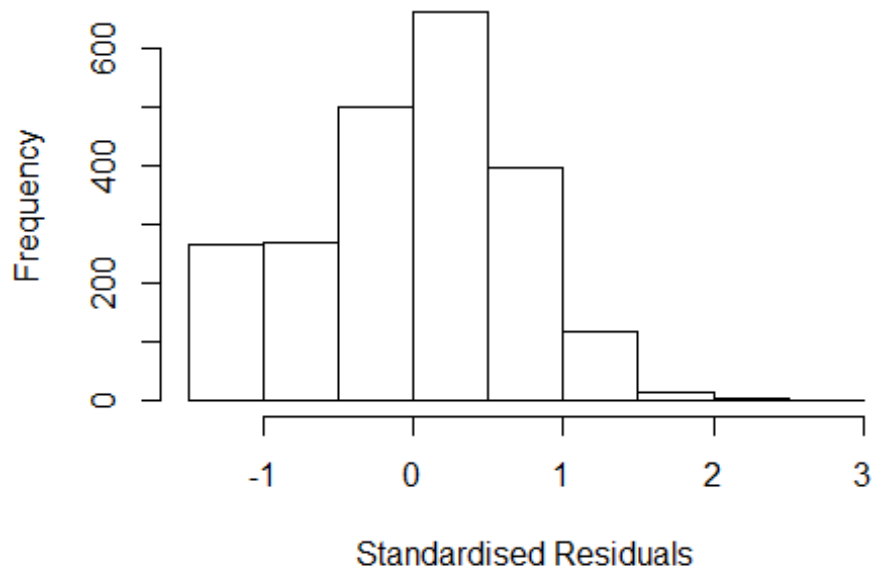
```

## Year2007          -0.1443      0.3086   -0.47      0.64
## Year2008          -0.0449      0.3092   -0.15      0.88
## Year2009          -0.1162      0.3100   -0.38      0.71
## Year2010           0.0207      0.3081    0.07      0.95
## Year2011          -0.0464      0.3088   -0.15      0.88
## Year2012          -0.0509      0.3066   -0.17      0.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.683
## Multiple R-squared:  0.00842,    Adjusted R-squared:  0.000813
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 190 weights are ~= 1. The remaining 2043 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.863  0.946  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.48e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1      1.016
## Year              1.033 16      1.001

```



## Residuals from last author



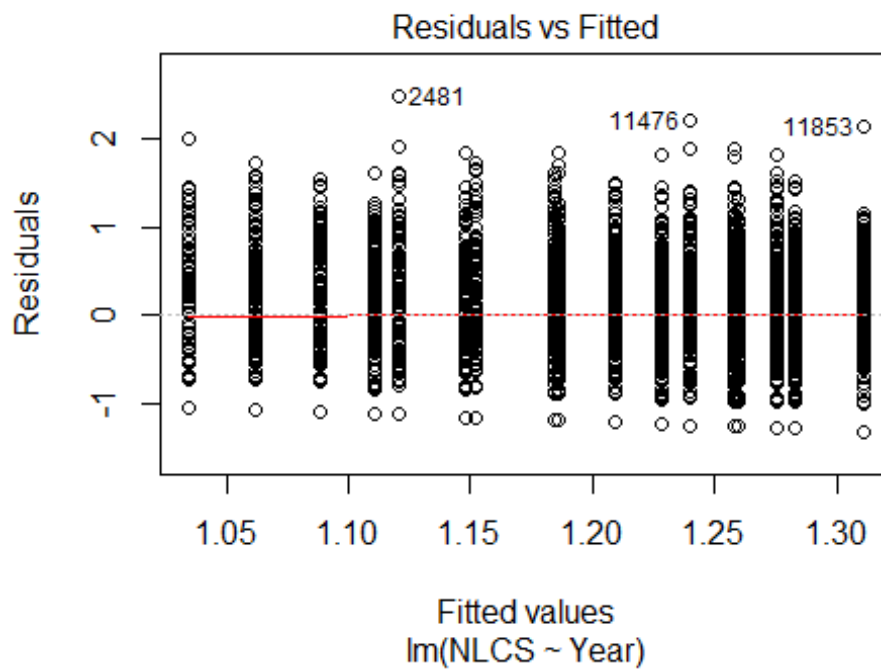
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 172 0031200344 3.606 1997    2102      2    2.629
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.244 -0.474  0.059  0.470  2.631
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18645    0.30506   3.89  0.0001 ***
## LastAuthorFemale1  0.00228    0.04183   0.05  0.9566
## Year1997      -0.21125    0.31272  -0.68  0.4994
## Year1998      -0.05065    0.31360  -0.16  0.8717
## Year1999      -0.14000    0.31528  -0.44  0.6570
## Year2000      -0.08638    0.31136  -0.28  0.7815
## Year2001      -0.07849    0.31481  -0.25  0.8031
## Year2002      -0.01391    0.31286  -0.04  0.9646
## Year2003      -0.04789    0.31075  -0.15  0.8775
## Year2004      -0.01406    0.31156  -0.05  0.9640
## Year2005       0.05497    0.31197   0.18  0.8602
## Year2006      -0.09307    0.31343  -0.30  0.7665
```

```

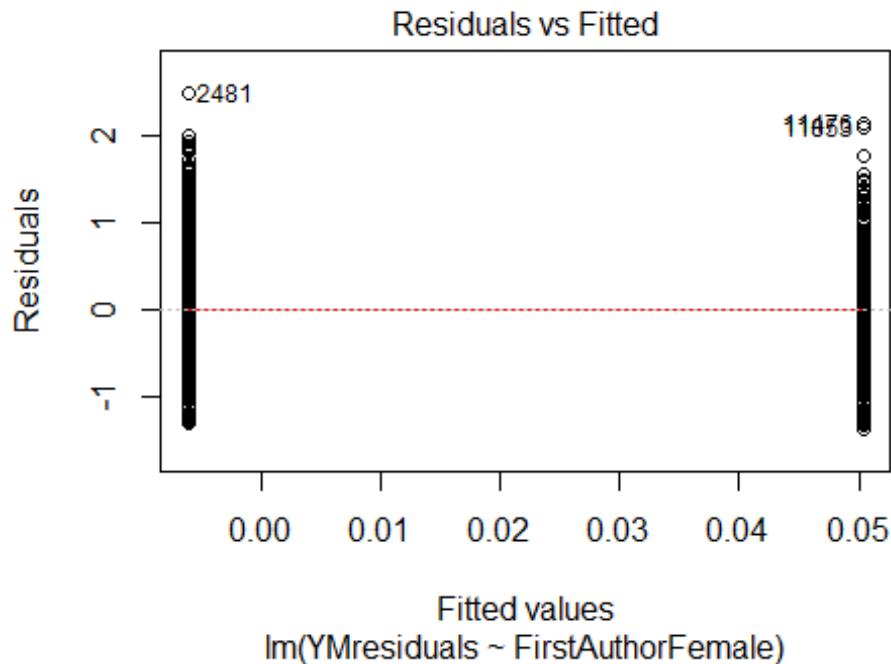
## Year2007          -0.14678      0.30952    -0.47    0.6354
## Year2008          -0.04652      0.31027    -0.15    0.8808
## Year2009          -0.11916      0.31094    -0.38    0.7016
## Year2010           0.01892      0.30917     0.06    0.9512
## Year2011          -0.04771      0.30985    -0.15    0.8776
## Year2012          -0.05262      0.30768    -0.17    0.8642
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.683
## Multiple R-squared:  0.00839,    Adjusted R-squared:  0.000779
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 191 weights are ~= 1. The remaining 2042 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.105  0.863  0.947   0.910  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.48e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2233"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2210"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1370 1374 1313  878 1069 1226 1033  993  961  981 1147 1177 1088 1168 1069
## 2011 2012
## 1087 1117
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  153  165  166  102  143  124  154  192  170  213  230  263  248  308  277
## 2011 2012

```

```
## 353 423
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 128 145 145 95 125 105 137 173 144 161 185 205 199 262 231
## 2011 2012
## 302 362
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 85, df = 16, p-value = 2e-11
```

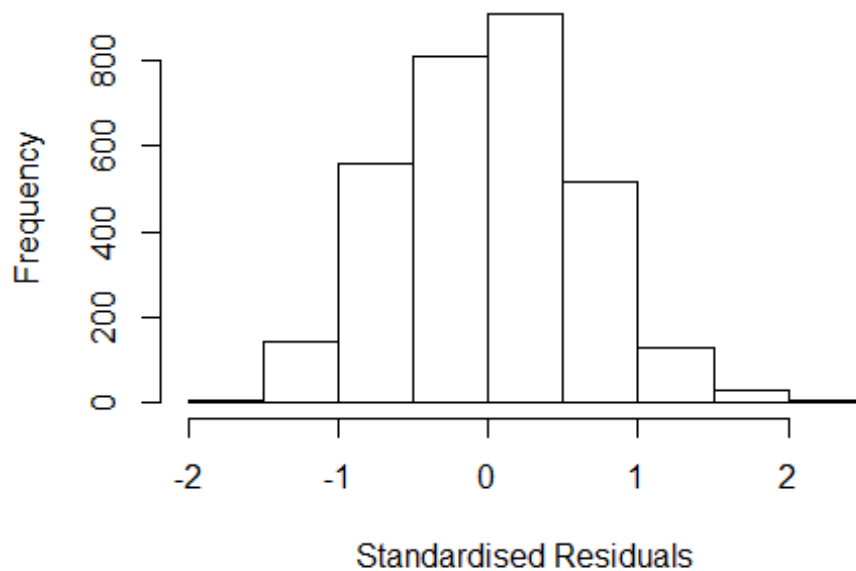


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.57, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 3.6056917576721"
## [1] "Male first author team size 2018 geometric mean: 3.03940155420302"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 25000, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.30630388888874"
## [1] "Male last author team size 2018 geometric mean: 3.10807068891619"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1      1.043
## LastAuthorFemale  1.072 1      1.035
## UniqueAuthors     1.250 4      1.028
## Year              1.242 16      1.007
```

## Residuals from first and last author and team size



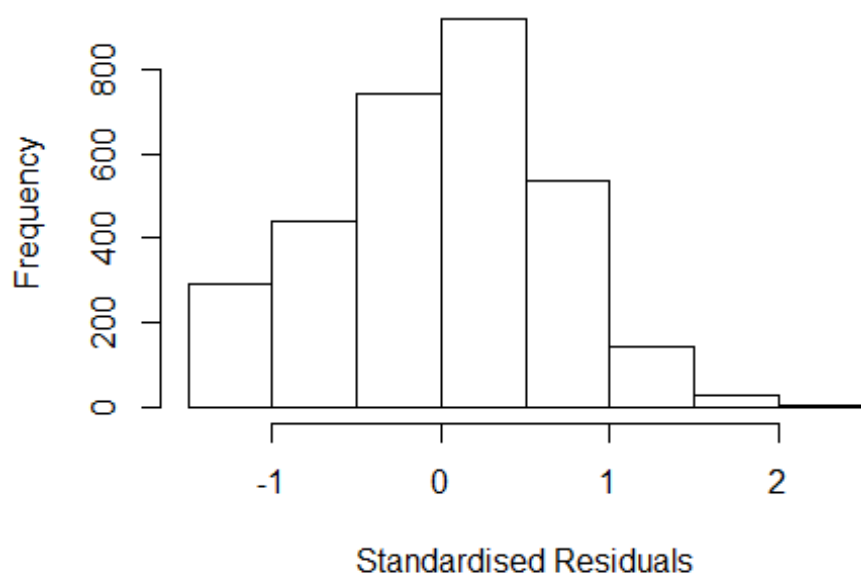
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5565 -0.4373 0.0177 0.4398 2.2892
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7351 0.0644 11.42 <2e-16 ***
## FirstAuthorFemale1 0.0215 0.0405 0.53 0.5956
## LastAuthorFemale1 -0.0120 0.0481 -0.25 0.8037
## UniqueAuthors2 0.4064 0.0331 12.28 <2e-16 ***
## UniqueAuthors3 0.4347 0.0354 12.29 <2e-16 ***
## UniqueAuthors4 0.5288 0.0412 12.84 <2e-16 ***
## UniqueAuthors5 0.5689 0.0454 12.52 <2e-16 ***
## Year1997 0.1340 0.0895 1.50 0.1345
## Year1998 0.1398 0.0942 1.48 0.1379
## Year1999 0.0235 0.1053 0.22 0.8238
```

```

## Year2000          0.0161      0.0892      0.18      0.8568
## Year2001          0.0897      0.1015      0.88      0.3768
## Year2002          0.1240      0.0875      1.42      0.1563
## Year2003          0.1418      0.0868      1.63      0.1023
## Year2004          0.1778      0.0856      2.08      0.0378 *
## Year2005          0.2489      0.0814      3.06      0.0022 **
## Year2006          0.2309      0.0806      2.87      0.0042 **
## Year2007          0.1507      0.0772      1.95      0.0511 .
## Year2008          0.0951      0.0783      1.21      0.2251
## Year2009          0.1845      0.0735      2.51      0.0120 *
## Year2010          0.1679      0.0741      2.27      0.0234 *
## Year2011          0.1900      0.0707      2.69      0.0072 **
## Year2012          0.0681      0.0705      0.97      0.3346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.647
## Multiple R-squared:  0.114, Adjusted R-squared:  0.107
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 283 weights are ~= 1. The remaining 2821 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.184  0.869  0.948  0.910  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.22e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1          1.029
## LastAuthorFemale  1.051 1          1.025
## Year              1.028 16          1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3471 -0.4610 0.0419 0.4707 2.4774
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.95658 0.07105 13.46 < 2e-16 ***
## FirstAuthorFemale1 0.06666 0.04169 1.60 0.10992
## LastAuthorFemale1 0.00515 0.04954 0.10 0.91718
## Year1997 0.15906 0.09994 1.59 0.11159
## Year1998 0.12997 0.10434 1.25 0.21299
## Year1999 0.03187 0.11819 0.27 0.78744
## Year2000 0.06624 0.10016 0.66 0.50843
## Year2001 0.16969 0.10701 1.59 0.11290
## Year2002 0.16267 0.09640 1.69 0.09162 .
## Year2003 0.21739 0.09510 2.29 0.02233 *
## Year2004 0.21802 0.09286 2.35 0.01894 *
## Year2005 0.36171 0.08892 4.07 4.9e-05 ***
```

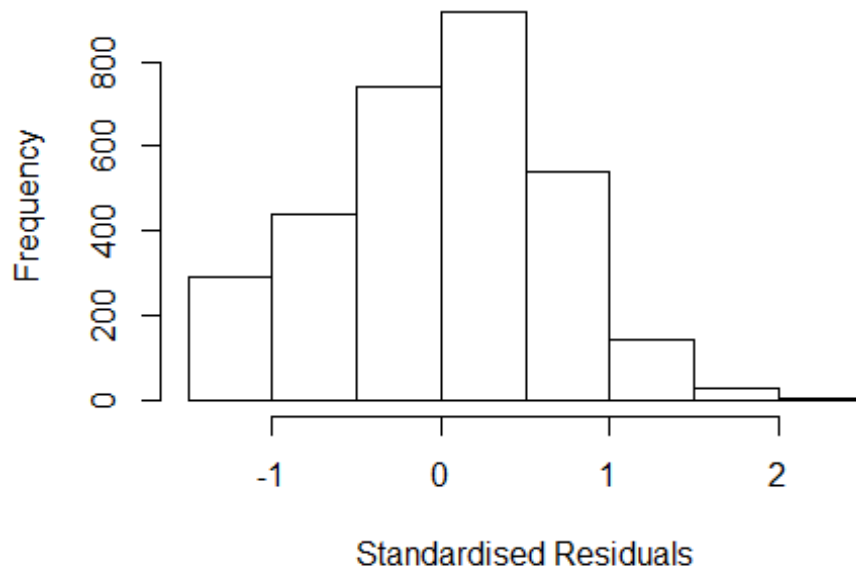
```

## Year2006          0.29085      0.08802      3.30  0.00096 ***
## Year2007          0.28721      0.08486      3.38  0.00072 ***
## Year2008          0.20756      0.08689      2.39  0.01696 *
## Year2009          0.31898      0.08085      3.95  8.2e-05 ***
## Year2010          0.31873      0.08060      3.95  7.8e-05 ***
## Year2011          0.32150      0.07864      4.09  4.5e-05 ***
## Year2012          0.24983      0.07757      3.22  0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.666
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.0148
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 254 weights are ~= 1. The remaining 2850 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.137  0.865   0.948   0.908   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.22e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.02 1      1.010
## Year              1.02 16      1.001

```



## Residuals from first author



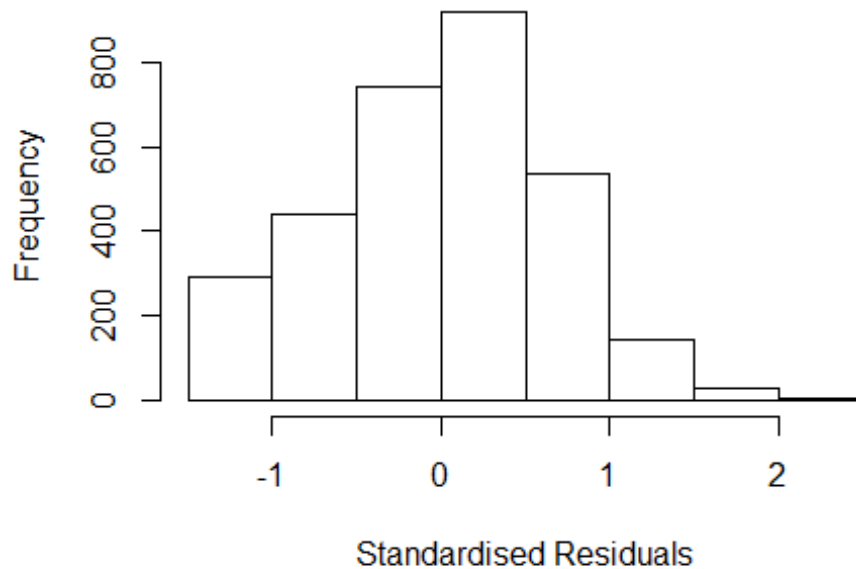
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3461 -0.4621 0.0425 0.4705 2.4771
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9567 0.0710 13.47 < 2e-16 ***
## FirstAuthorFemale1 0.0677 0.0410 1.65 0.09822 .
## Year1997 0.1592 0.0999 1.59 0.11139
## Year1998 0.1301 0.1044 1.25 0.21259
## Year1999 0.0320 0.1182 0.27 0.78683
## Year2000 0.0664 0.1002 0.66 0.50721
## Year2001 0.1698 0.1070 1.59 0.11261
## Year2002 0.1628 0.0964 1.69 0.09148 .
## Year2003 0.2176 0.0951 2.29 0.02215 *
## Year2004 0.2183 0.0929 2.35 0.01881 *
## Year2005 0.3619 0.0889 4.07 4.8e-05 ***
## Year2006 0.2912 0.0880 3.31 0.00094 ***
```

```

## Year2007          0.2874      0.0848      3.39  0.00071 ***
## Year2008          0.2079      0.0869      2.39  0.01678 *
## Year2009          0.3192      0.0808      3.95  8.0e-05 ***
## Year2010          0.3189      0.0806      3.96  7.8e-05 ***
## Year2011          0.3217      0.0786      4.09  4.4e-05 ***
## Year2012          0.2499      0.0776      3.22  0.00129 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.666
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.0151
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 2849 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.137  0.865  0.948  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.22e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1      1.006
## Year      1.012 16      1.000

```

## Residuals from last author



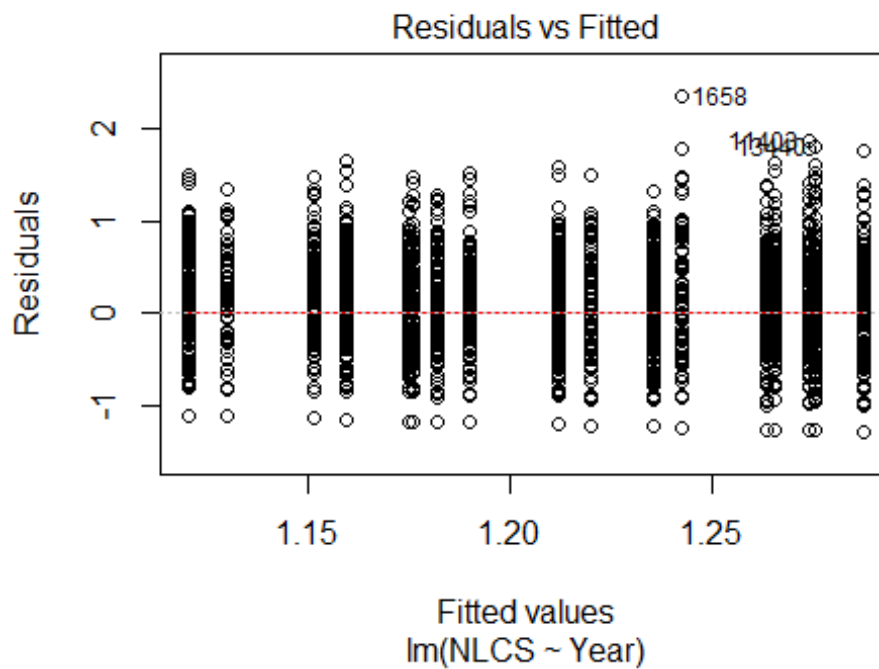
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3454 -0.4593 0.0392 0.4665 2.4734
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9586 0.0712 13.46 < 2e-16 ***
## LastAuthorFemale1 0.0231 0.0483 0.48 0.63247
## Year1997 0.1610 0.1000 1.61 0.10753
## Year1998 0.1323 0.1042 1.27 0.20417
## Year1999 0.0343 0.1186 0.29 0.77263
## Year2000 0.0720 0.1006 0.72 0.47406
## Year2001 0.1692 0.1072 1.58 0.11457
## Year2002 0.1657 0.0965 1.72 0.08621 .
## Year2003 0.2211 0.0950 2.33 0.01997 *
## Year2004 0.2208 0.0928 2.38 0.01743 *
## Year2005 0.3638 0.0890 4.09 4.5e-05 ***
## Year2006 0.2970 0.0880 3.38 0.00075 ***
```

```

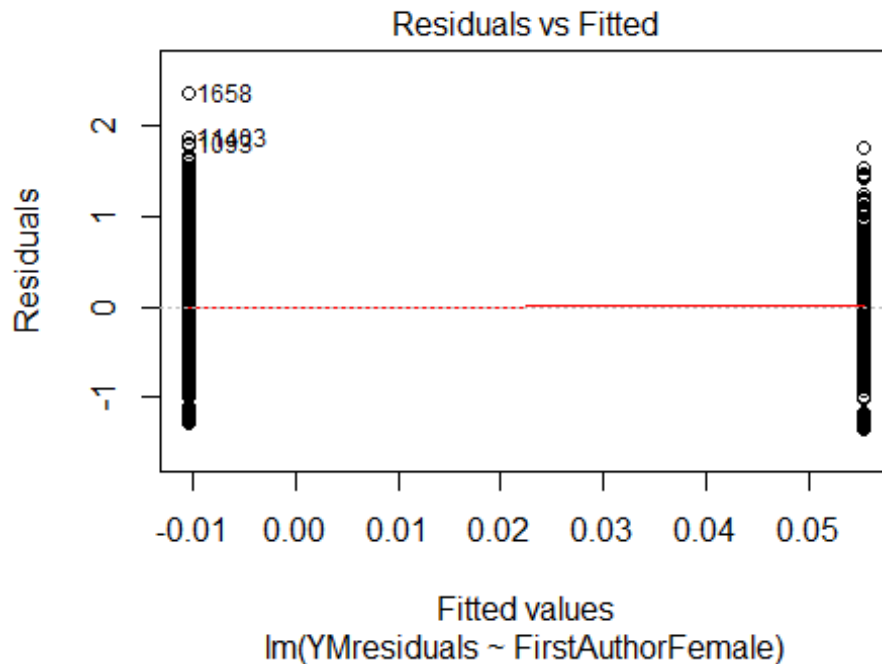
## Year2007          0.2919      0.0850      3.43  0.00060 ***
## Year2008          0.2114      0.0869      2.43  0.01500 *
## Year2009          0.3224      0.0809      3.98  6.9e-05 ***
## Year2010          0.3221      0.0807      3.99  6.7e-05 ***
## Year2011          0.3252      0.0787      4.13  3.7e-05 ***
## Year2012          0.2552      0.0776      3.29  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.666
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0142
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 248 weights are ~= 1. The remaining 2856 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.866  0.949  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.22e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3104"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2211"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  963  928  965  551  701  908  782  728  799  729  716  727  691  841  768
## 2011 2012
##  827  789
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  126  112  141   73  100  107  132  142  161  181  192  177  189  246  230
## 2011 2012

```

```
## 259 269
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 89 119 68 89 91 116 127 146 140 162 146 153 211 196
## 2011 2012
## 226 236
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 54, df = 16, p-value = 5e-06
```

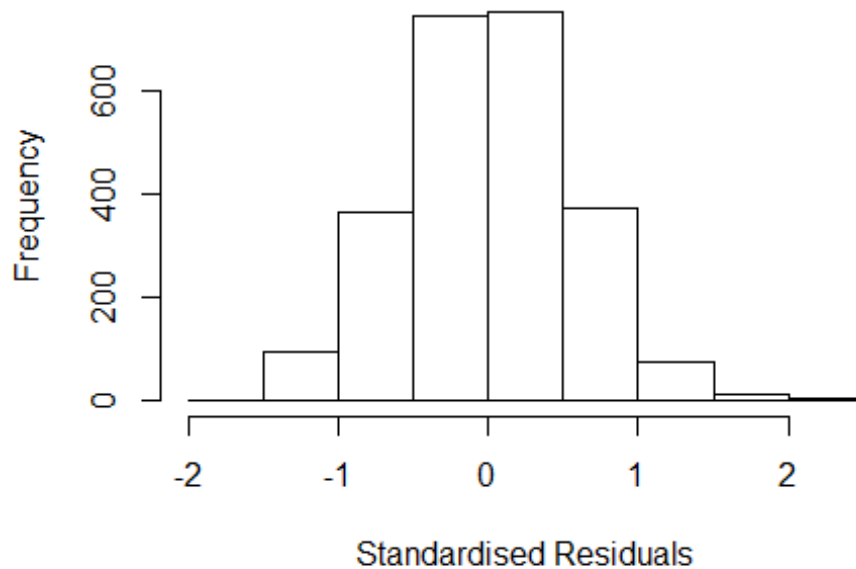


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.62787485836618"
## [1] "Male first author team size 2018 geometric mean: 2.89856759865588"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8400, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.26420581876379"
## [1] "Male last author team size 2018 geometric mean: 2.97568581606777"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.131 1      1.063
## LastAuthorFemale  1.116 1      1.057
## UniqueAuthors     1.283 4      1.032
## Year              1.270 16     1.008
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.50676 -0.36655  0.00279  0.38090  2.21684
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8965    0.0630   14.24  <2e-16 ***
## FirstAuthorFemale1  0.0320    0.0318    1.01   0.314
## LastAuthorFemale1 -0.0143    0.0415   -0.34   0.731
## UniqueAuthors2    0.2989    0.0357    8.38  <2e-16 ***
## UniqueAuthors3    0.3479    0.0359    9.70  <2e-16 ***
## UniqueAuthors4    0.3583    0.0409    8.76  <2e-16 ***
## UniqueAuthors5    0.4220    0.0422   10.01  <2e-16 ***
## Year1997          0.1318    0.0960    1.37   0.170
## Year1998          0.0496    0.0889    0.56   0.577
## Year1999          0.0522    0.0990    0.53   0.598
```

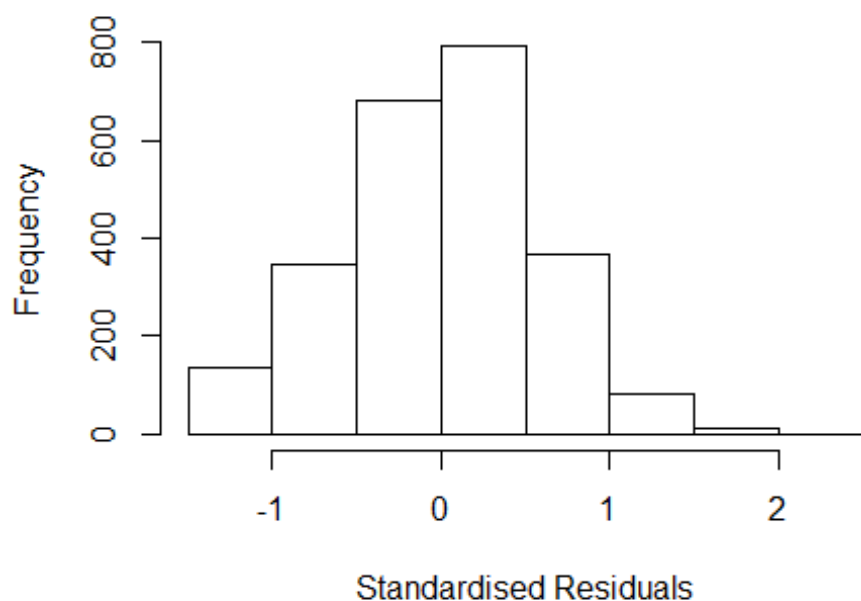
```

## Year2000          0.0394      0.0964      0.41      0.683
## Year2001          0.0669      0.0955      0.70      0.484
## Year2002          0.0600      0.0828      0.72      0.469
## Year2003          0.0716      0.0829      0.86      0.388
## Year2004          0.1162      0.0788      1.47      0.140
## Year2005          0.0288      0.0802      0.36      0.720
## Year2006          0.1563      0.0747      2.09      0.037 *
## Year2007          0.0803      0.0780      1.03      0.304
## Year2008          0.0907      0.0760      1.19      0.233
## Year2009          0.0934      0.0712      1.31      0.190
## Year2010          0.0616      0.0703      0.88      0.381
## Year2011          0.0207      0.0698      0.30      0.767
## Year2012         -0.0438      0.0704     -0.62      0.534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0773, Adjusted R-squared:  0.0689
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 232 weights are ~= 1. The remaining 2189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0801 0.8730 0.9470 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.077 1      1.038
## LastAuthorFemale  1.089 1      1.043
## Year              1.055 16      1.002

```



## Residuals from first and last author

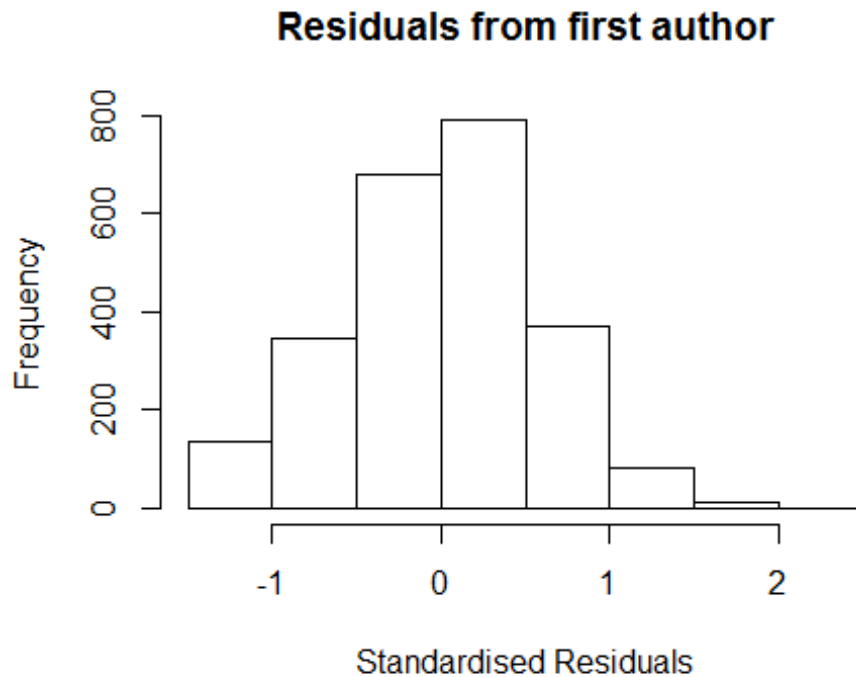


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3767 -0.3884 0.0319 0.3910 2.3585
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07832 0.06586 16.37 <2e-16 ***
## FirstAuthorFemale1 0.09608 0.03192 3.01 0.0026 **
## LastAuthorFemale1 -0.00346 0.04276 -0.08 0.9356
## Year1997 0.15616 0.10105 1.55 0.1224
## Year1998 0.07209 0.09385 0.77 0.4425
## Year1999 0.05522 0.10830 0.51 0.6101
## Year2000 0.10406 0.10443 1.00 0.3191
## Year2001 0.16206 0.10031 1.62 0.1063
## Year2002 0.09134 0.08618 1.06 0.2893
## Year2003 0.12491 0.08908 1.40 0.1610
## Year2004 0.16214 0.08436 1.92 0.0547 .
## Year2005 0.10730 0.08449 1.27 0.2042
```

```

## Year2006          0.20232      0.07942      2.55      0.0109 *
## Year2007          0.16670      0.08261      2.02      0.0437 *
## Year2008          0.17256      0.08028      2.15      0.0317 *
## Year2009          0.16692      0.07638      2.19      0.0290 *
## Year2010          0.16479      0.07433      2.22      0.0267 *
## Year2011          0.10251      0.07440      1.38      0.1684
## Year2012          0.05978      0.07500      0.80      0.4255
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.0116, Adjusted R-squared:  0.00422
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 201 weights are ~= 1. The remaining 2220 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0509 0.8690 0.9500 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.011
## Year              1.023 16      1.001

```

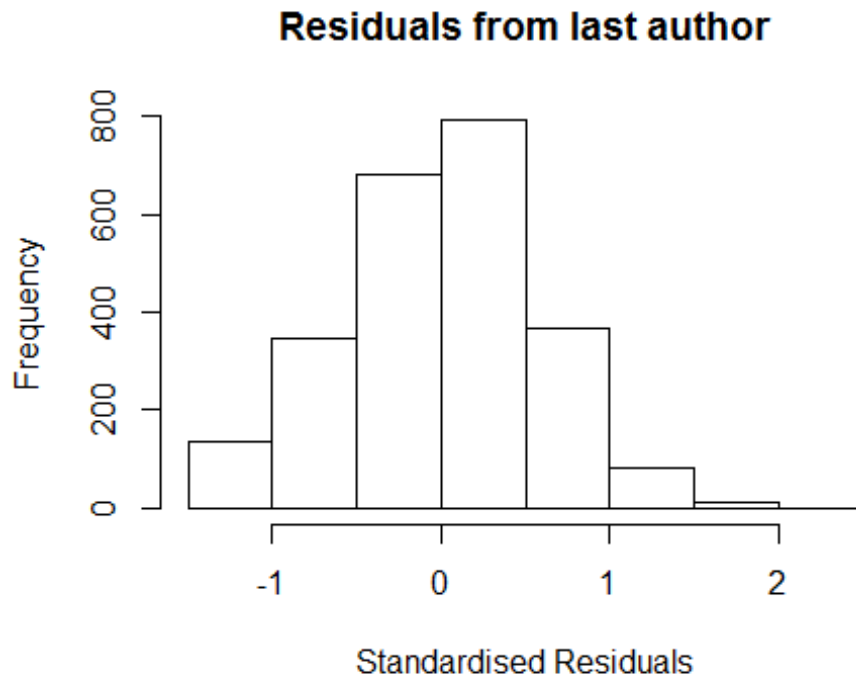


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3757 -0.3890 0.0323 0.3912 2.3585
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0782 0.0658 16.38 <2e-16 ***
## FirstAuthorFemale1 0.0954 0.0311 3.07 0.0022 **
## Year1997 0.1563 0.1010 1.55 0.1220
## Year1998 0.0721 0.0938 0.77 0.4421
## Year1999 0.0553 0.1082 0.51 0.6094
## Year2000 0.1041 0.1044 1.00 0.3189
## Year2001 0.1620 0.1003 1.62 0.1064
## Year2002 0.0913 0.0862 1.06 0.2894
## Year2003 0.1248 0.0891 1.40 0.1614
## Year2004 0.1623 0.0843 1.92 0.0544 .
## Year2005 0.1073 0.0845 1.27 0.2043
## Year2006 0.2022 0.0794 2.55 0.0110 *
```

```

## Year2007          0.1666      0.0826      2.02      0.0437 *
## Year2008          0.1724      0.0803      2.15      0.0320 *
## Year2009          0.1669      0.0764      2.19      0.0289 *
## Year2010          0.1646      0.0743      2.22      0.0268 *
## Year2011          0.1023      0.0743      1.38      0.1687
## Year2012          0.0596      0.0750      0.80      0.4263
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.0116, Adjusted R-squared:  0.00463
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 204 weights are ~= 1. The remaining 2217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0514 0.8690 0.9500 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1      1.016
## Year      1.033 16      1.001

```



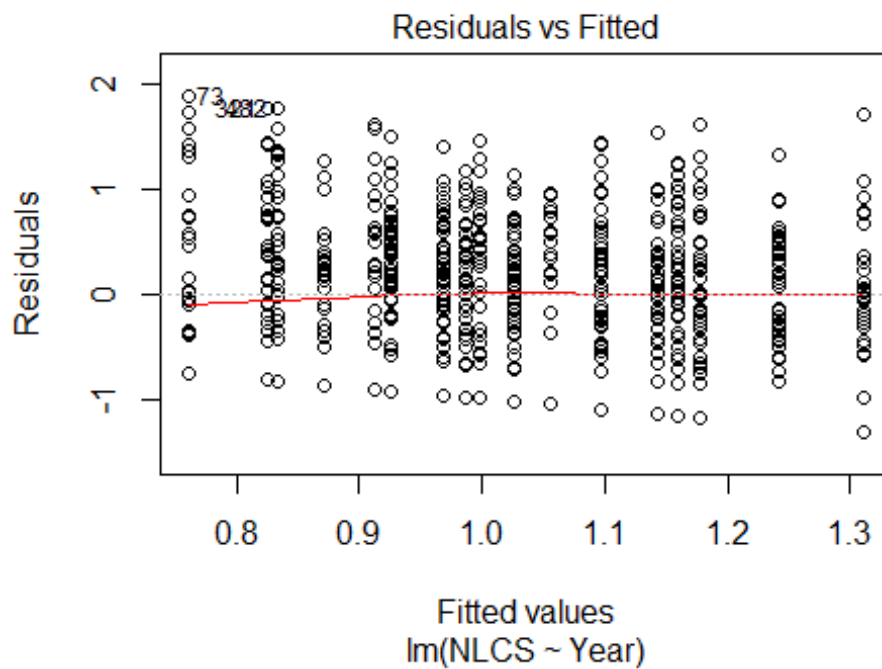
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3224 -0.3853 0.0314 0.3898 2.3470
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0847 0.0659 16.45 <2e-16 ***
## LastAuthorFemale1 0.0261 0.0416 0.63 0.5296
## Year1997 0.1613 0.1008 1.60 0.1097
## Year1998 0.0780 0.0937 0.83 0.4054
## Year1999 0.0592 0.1088 0.54 0.5867
## Year2000 0.1115 0.1054 1.06 0.2902
## Year2001 0.1615 0.1005 1.61 0.1082
## Year2002 0.0939 0.0868 1.08 0.2793
## Year2003 0.1295 0.0889 1.46 0.1452
## Year2004 0.1728 0.0846 2.04 0.0411 *
## Year2005 0.1213 0.0845 1.44 0.1511
## Year2006 0.2116 0.0795 2.66 0.0078 **
```

```

## Year2007          0.1786      0.0828      2.16      0.0311 *
## Year2008          0.1784      0.0803      2.22      0.0263 *
## Year2009          0.1754      0.0765      2.29      0.0219 *
## Year2010          0.1696      0.0744      2.28      0.0227 *
## Year2011          0.1109      0.0744      1.49      0.1363
## Year2012          0.0649      0.0751      0.86      0.3876
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.00821,    Adjusted R-squared:  0.00119
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 197 weights are ~= 1. The remaining 2224 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0551 0.8680 0.9510 0.9030 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2421"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2212"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 133 147 165 136 133 117 102 111 107 99 115 115 104 125 107
## 2011 2012
## 120 119
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 45 55 32 33 26 41 45 49 44 40 52 35 43 51
## 2011 2012

```

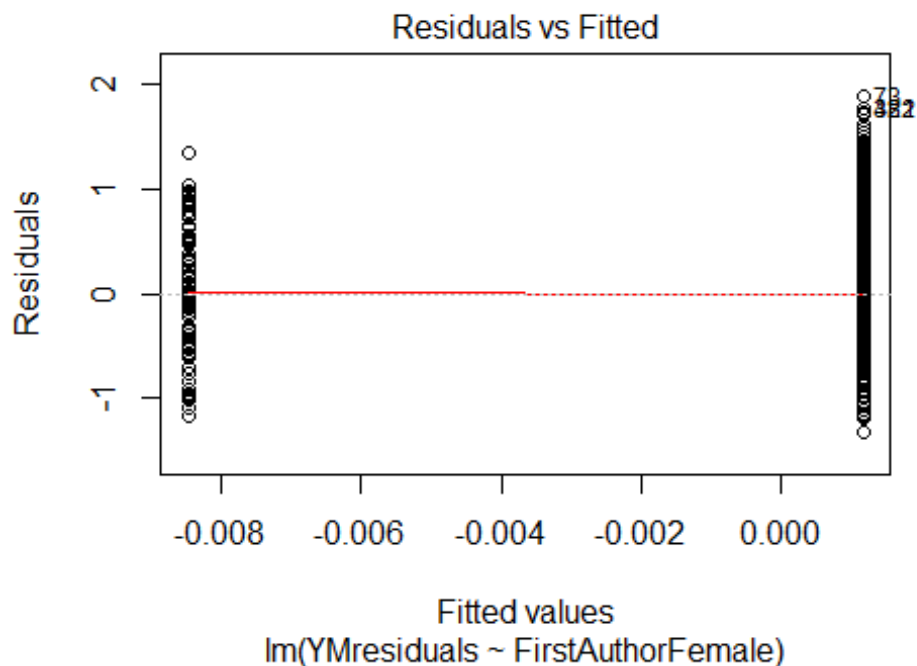
```
## 56 58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 44 54 29 28 23 39 38 41 43 37 38 31 39 42
## 2011 2012
## 49 54
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 18, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.5, df = 1, p-value = 0.03
## [1] "Female first author team size 2018 geometric mean: 2.74674363166598"
## [1] "Male first author team size 2018 geometric mean: 3.22920898550094"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.70213350320354"
## [1] "Male last author team size 2018 geometric mean: 3.25147299236692"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

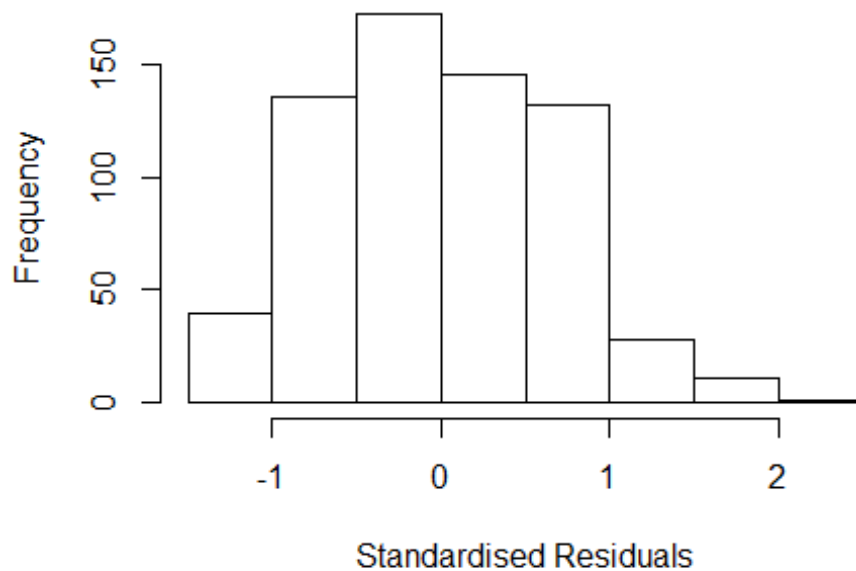


```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.312	1	1.145
LastAuthorFemale	1.274	1	1.129
UniqueAuthors	1.650	4	1.065
Year	1.859	16	1.020



## Residuals from first and last author and team size



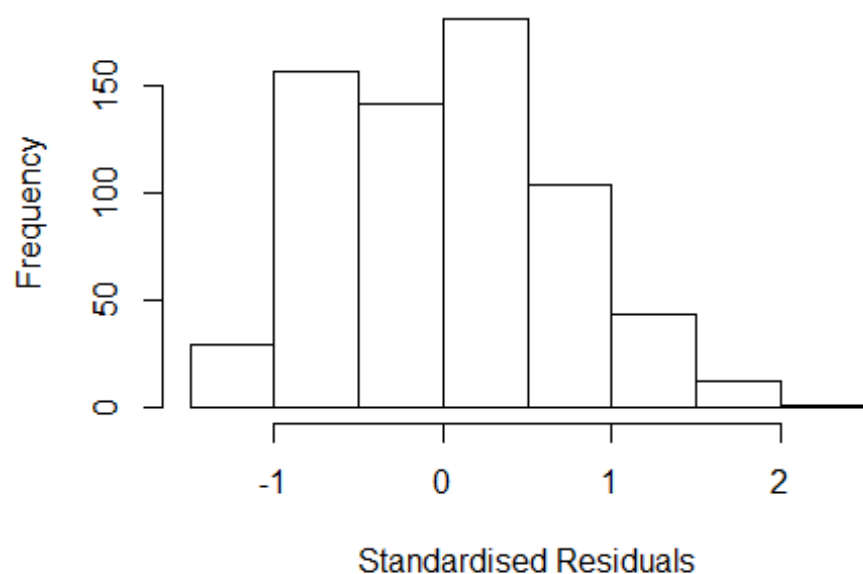
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3427 -0.5515 -0.0316 0.5181 2.1914
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4476 0.1084 4.13 4.1e-05 ***
## FirstAuthorFemale1 -0.0201 0.0758 -0.26 0.7915
## LastAuthorFemale1 -0.0105 0.1006 -0.10 0.9172
## UniqueAuthors2 0.4330 0.0710 6.10 1.9e-09 ***
## UniqueAuthors3 0.5794 0.0823 7.04 4.9e-12 ***
## UniqueAuthors4 0.5301 0.1072 4.94 9.8e-07 ***
## UniqueAuthors5 0.6414 0.1125 5.70 1.8e-08 ***
## Year1997 0.1784 0.1606 1.11 0.2671
## Year1998 0.1545 0.1427 1.08 0.2794
## Year1999 0.2104 0.1876 1.12 0.2625
```

```

## Year2000          0.1673      0.1784      0.94      0.3488
## Year2001          0.3499      0.1748      2.00      0.0457 *
## Year2002          0.2373      0.1727      1.37      0.1700
## Year2003          0.3180      0.1553      2.05      0.0411 *
## Year2004          0.3262      0.1575      2.07      0.0388 *
## Year2005          0.1923      0.1548      1.24      0.2145
## Year2006          0.3874      0.1497      2.59      0.0099 **
## Year2007          0.1363      0.1443      0.94      0.3452
## Year2008          0.4464      0.1551      2.88      0.0041 **
## Year2009          0.3832      0.1367      2.80      0.0052 **
## Year2010          0.2251      0.1529      1.47      0.1416
## Year2011          0.1486      0.1366      1.09      0.2773
## Year2012          0.1646      0.1447      1.14      0.2560
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.686
## Multiple R-squared:  0.162, Adjusted R-squared:  0.134
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 612 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.286  0.890  0.939   0.916  0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.50e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.206 1      1.098
## LastAuthorFemale  1.234 1      1.111
## Year              1.230 16      1.006

```

## Residuals from first and last author



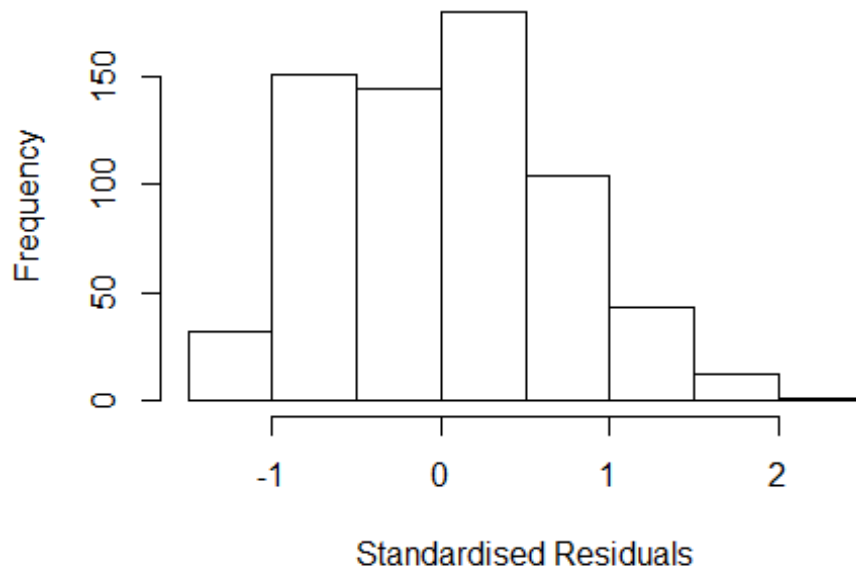
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.317 -0.569 0.031 0.467 2.051
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5881 0.1369 4.29 2.0e-05 ***
## FirstAuthorFemale1 0.0380 0.0778 0.49 0.62489
## LastAuthorFemale1 -0.0374 0.1051 -0.36 0.72209
## Year1997 0.2157 0.1886 1.14 0.25312
## Year1998 0.1776 0.1747 1.02 0.30957
## Year1999 0.2468 0.2173 1.14 0.25655
## Year2000 0.3034 0.1829 1.66 0.09757 .
## Year2001 0.4675 0.2120 2.20 0.02783 *
## Year2002 0.4006 0.1955 2.05 0.04082 *
## Year2003 0.5098 0.1850 2.76 0.00603 **
## Year2004 0.4474 0.1790 2.50 0.01269 *
## Year2005 0.3898 0.1771 2.20 0.02809 *
```

```

## Year2006          0.5732      0.1808      3.17  0.00159 **
## Year2007          0.3307      0.1711      1.93  0.05365 .
## Year2008          0.7290      0.1827      3.99  7.4e-05 ***
## Year2009          0.5901      0.1603      3.68  0.00025 ***
## Year2010          0.4700      0.1710      2.75  0.00614 **
## Year2011          0.3797      0.1654      2.30  0.02201 *
## Year2012          0.3141      0.1740      1.81  0.07152 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.726
## Multiple R-squared:  0.0512, Adjusted R-squared:  0.0249
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 625 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.406  0.867  0.944  0.917  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.099 1          1.048
## Year              1.099 16          1.003

```

## Residuals from first author



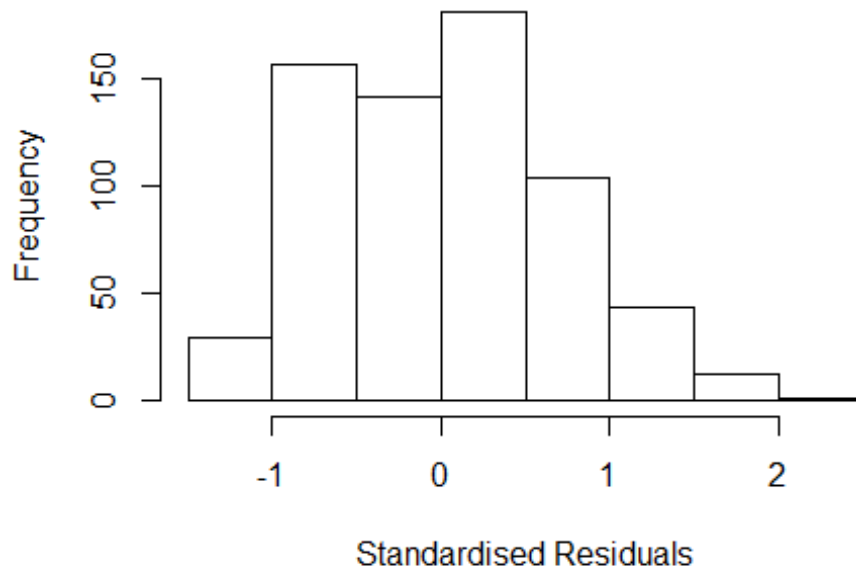
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3152 -0.5723 0.0328 0.4673 2.0543
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5847 0.1356 4.31 1.9e-05 ***
## FirstAuthorFemale1 0.0274 0.0750 0.37 0.71513
## Year1997 0.2196 0.1871 1.17 0.24077
## Year1998 0.1795 0.1740 1.03 0.30245
## Year1999 0.2481 0.2173 1.14 0.25404
## Year2000 0.3025 0.1825 1.66 0.09788 .
## Year2001 0.4694 0.2112 2.22 0.02663 *
## Year2002 0.4013 0.1955 2.05 0.04044 *
## Year2003 0.5128 0.1836 2.79 0.00537 **
## Year2004 0.4449 0.1789 2.49 0.01315 *
## Year2005 0.3901 0.1770 2.20 0.02789 *
## Year2006 0.5740 0.1800 3.19 0.00150 **
```

```

## Year2007          0.3333      0.1703      1.96  0.05085 .
## Year2008          0.7305      0.1821      4.01  6.7e-05 ***
## Year2009          0.5938      0.1594      3.73  0.00021 ***
## Year2010          0.4713      0.1703      2.77  0.00581 **
## Year2011          0.3811      0.1645      2.32  0.02086 *
## Year2012          0.3137      0.1736      1.81  0.07118 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.728
## Multiple R-squared:  0.051, Adjusted R-squared:  0.0262
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 628 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.406  0.869  0.945  0.917  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.123 1          1.060
## Year            1.123 16          1.004

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.322 -0.566 0.042 0.463 2.050
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5889 0.1374 4.28 2.1e-05 ***
## LastAuthorFemale1 -0.0232 0.0998 -0.23 0.81598
## Year1997 0.2165 0.1888 1.15 0.25186
## Year1998 0.1771 0.1750 1.01 0.31182
## Year1999 0.2475 0.2182 1.13 0.25710
## Year2000 0.3065 0.1825 1.68 0.09349 .
## Year2001 0.4678 0.2125 2.20 0.02805 *
## Year2002 0.4040 0.1949 2.07 0.03857 *
## Year2003 0.5136 0.1841 2.79 0.00544 **
## Year2004 0.4507 0.1790 2.52 0.01207 *
## Year2005 0.3937 0.1770 2.22 0.02644 *
## Year2006 0.5782 0.1793 3.22 0.00133 **
```

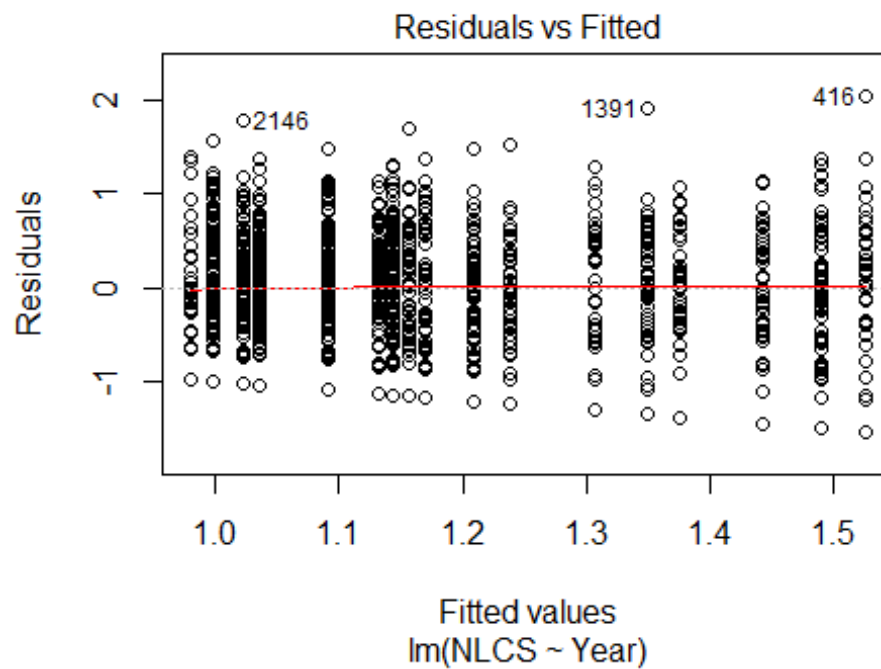
```

## Year2007          0.3364      0.1696      1.98  0.04772 *
## Year2008          0.7334      0.1820      4.03  6.3e-05 ***
## Year2009          0.5936      0.1601      3.71  0.00023 ***
## Year2010          0.4714      0.1709      2.76  0.00597 **
## Year2011          0.3815      0.1656      2.30  0.02153 *
## Year2012          0.3190      0.1731      1.84  0.06584 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.726
## Multiple R-squared:  0.0511, Adjusted R-squared:  0.0262
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 626 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.406  0.866  0.944  0.917  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 667"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2213"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 158 141 142 157 170 157 130 122 108 135 161 190 214 241 226
## 2011 2012
## 209 244
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 41 43 39 59 41 49 59 48 58 75 90 128 130 125
## 2011 2012

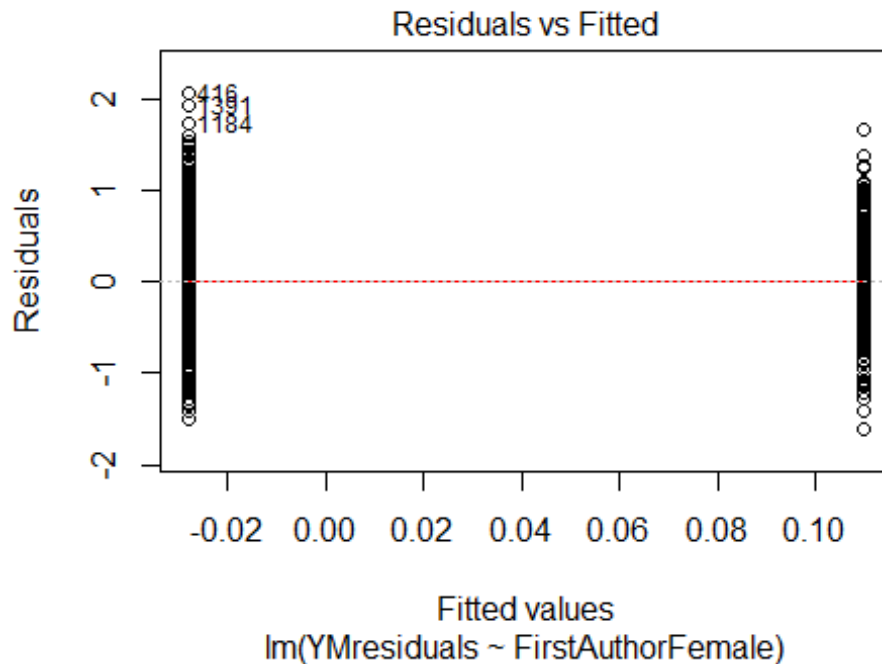
```



```
## 126 153
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 38 37 37 53 39 49 53 42 54 69 82 115 121 106
## 2011 2012
## 116 136
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.09
```

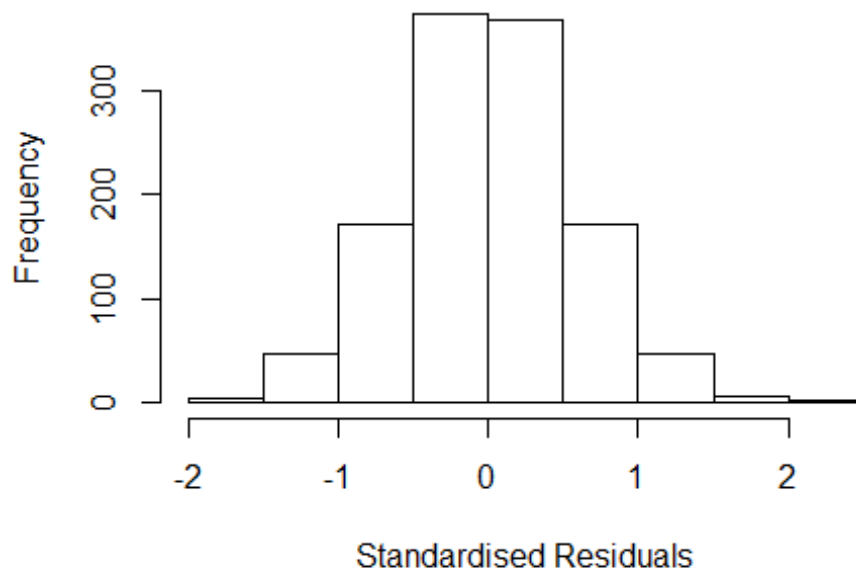


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.86, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.707943021874"
## [1] "Male first author team size 2018 geometric mean: 2.41542243828225"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3600, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.43958795343482"
## [1] "Male last author team size 2018 geometric mean: 2.52268305906813"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.291 1          1.136
## LastAuthorFemale  1.220 1          1.105
## UniqueAuthors     1.309 4          1.034
## Year              1.411 16          1.011
```

## Residuals from first and last author and team size



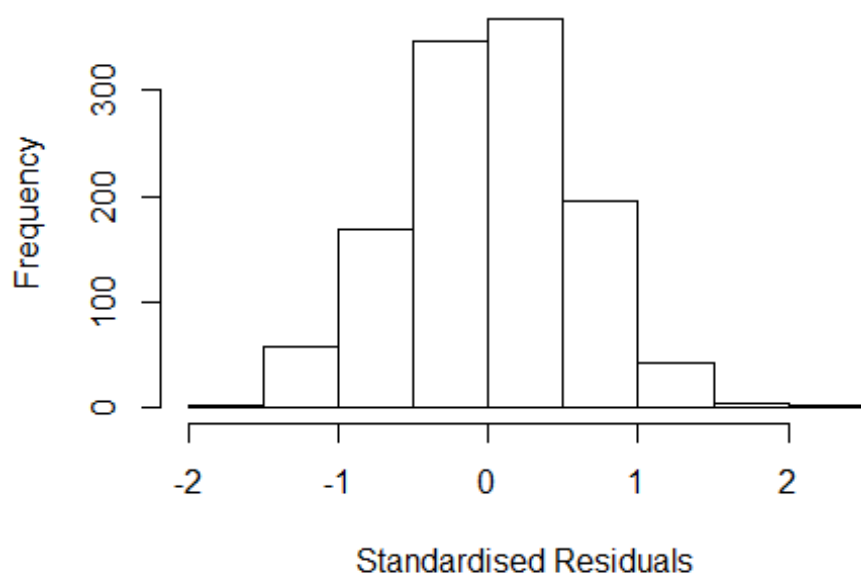
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.68584 -0.39096 -0.00234 0.39703 2.17400
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.285336 0.093614 13.73 < 2e-16 ***
## FirstAuthorFemale1 0.106775 0.045381 2.35 0.01880 *
## LastAuthorFemale1 -0.000286 0.049043 -0.01 0.99536
## UniqueAuthors2 0.202374 0.043493 4.65 3.6e-06 ***
## UniqueAuthors3 0.276450 0.052246 5.29 1.5e-07 ***
## UniqueAuthors4 0.286742 0.064430 4.45 9.4e-06 ***
## UniqueAuthors5 0.259470 0.073539 3.53 0.00043 ***
## Year1997 -0.018532 0.171634 -0.11 0.91404
## Year1998 0.097662 0.177177 0.55 0.58159
## Year1999 -0.488108 0.134467 -3.63 0.00030 ***
```

```

## Year2000      -0.270819    0.122393    -2.21    0.02711 *
## Year2001      -0.003266    0.130743    -0.02    0.98008
## Year2002      -0.276057    0.125604    -2.20    0.02816 *
## Year2003       0.017569    0.126695     0.14    0.88973
## Year2004      -0.091508    0.127542    -0.72    0.47323
## Year2005      -0.196480    0.114746    -1.71    0.08711 .
## Year2006      -0.166139    0.109102    -1.52    0.12808
## Year2007      -0.326175    0.110760    -2.94    0.00329 **
## Year2008      -0.403799    0.105398    -3.83    0.00013 ***
## Year2009      -0.465995    0.106720    -4.37    1.4e-05 ***
## Year2010      -0.320246    0.111930    -2.86    0.00430 **
## Year2011      -0.420899    0.104758    -4.02    6.3e-05 ***
## Year2012      -0.387359    0.104552    -3.70    0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.117, Adjusted R-squared:  0.1
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1090 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.115  0.873  0.949  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.42e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.265 1          1.125
## LastAuthorFemale  1.233 1          1.110
## Year              1.117 16          1.003

```

## Residuals from first and last author



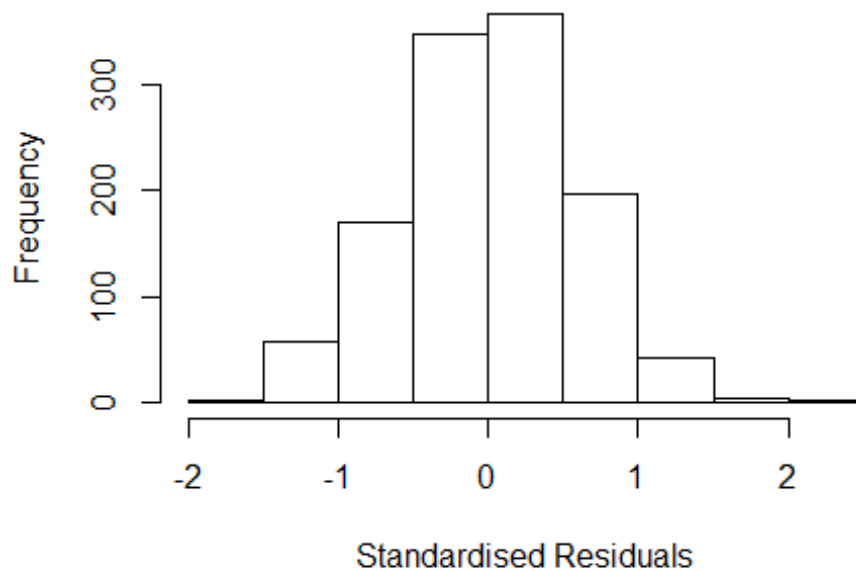
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5666 -0.4012 0.0119 0.3997 2.0644
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37434 0.09886 13.90 < 2e-16 ***
## FirstAuthorFemale1 0.15857 0.04605 3.44 0.00059 ***
## LastAuthorFemale1 -0.00951 0.05150 -0.18 0.85351
## Year1997 0.00242 0.16870 0.01 0.98857
## Year1998 0.11822 0.16524 0.72 0.47448
## Year1999 -0.46127 0.14049 -3.28 0.00106 **
## Year2000 -0.26035 0.13097 -1.99 0.04707 *
## Year2001 -0.00191 0.13644 -0.01 0.98881
## Year2002 -0.26395 0.13457 -1.96 0.05007 .
## Year2003 0.04318 0.13720 0.31 0.75300
## Year2004 -0.07396 0.13050 -0.57 0.57100
## Year2005 -0.14920 0.12106 -1.23 0.21806
```

```

## Year2006          -0.16536      0.11758      -1.41      0.15988
## Year2007          -0.27525      0.11613      -2.37      0.01794 *
## Year2008          -0.38163      0.11217      -3.40      0.00069 ***
## Year2009          -0.42383      0.11221      -3.78      0.00017 ***
## Year2010          -0.29093      0.11616      -2.50      0.01240 *
## Year2011          -0.38156      0.11138      -3.43      0.00063 ***
## Year2012          -0.34648      0.11114      -3.12      0.00187 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.0764, Adjusted R-squared:  0.0621
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1081 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.201  0.875   0.950   0.909   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1          1.032
## Year              1.065 16          1.002

```

## Residuals from first author



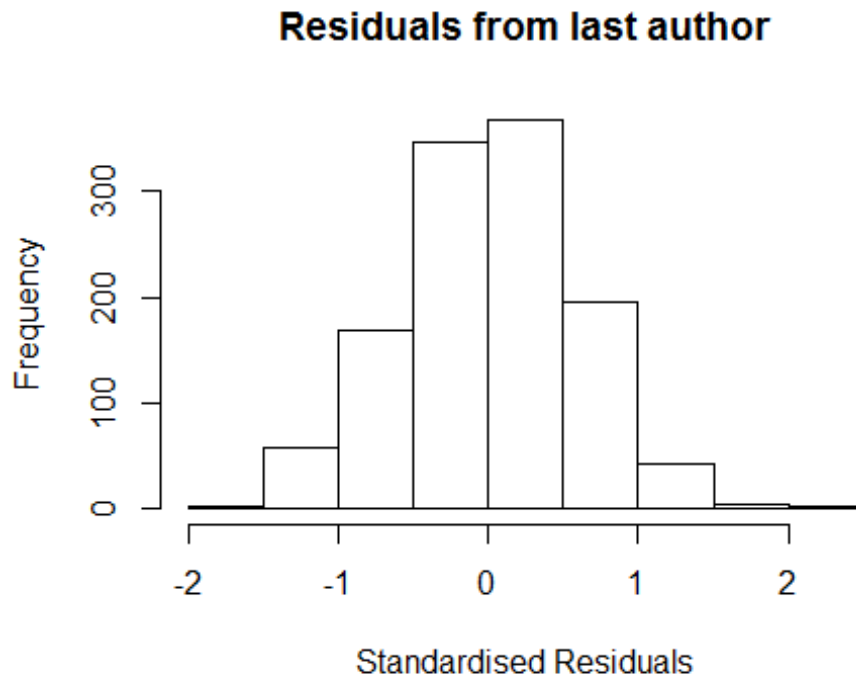
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.571 -0.401 0.013 0.401 2.065
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37444 0.09884 13.91 < 2e-16 ***
## FirstAuthorFemale1 0.15423 0.04259 3.62 0.00031 ***
## Year1997 0.00139 0.16867 0.01 0.99344
## Year1998 0.11745 0.16509 0.71 0.47696
## Year1999 -0.46152 0.14027 -3.29 0.00103 **
## Year2000 -0.26134 0.13073 -2.00 0.04583 *
## Year2001 -0.00338 0.13630 -0.02 0.98022
## Year2002 -0.26499 0.13421 -1.97 0.04856 *
## Year2003 0.04190 0.13707 0.31 0.75987
## Year2004 -0.07449 0.13041 -0.57 0.56794
## Year2005 -0.15003 0.12099 -1.24 0.21524
## Year2006 -0.16611 0.11757 -1.41 0.15797
```

```

## Year2007          -0.27643    0.11586   -2.39  0.01720 *
## Year2008          -0.38216    0.11219   -3.41  0.00068 ***
## Year2009          -0.42507    0.11168   -3.81  0.00015 ***
## Year2010          -0.29181    0.11599   -2.52  0.01201 *
## Year2011          -0.38214    0.11122   -3.44  0.00061 ***
## Year2012          -0.34699    0.11109   -3.12  0.00183 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.0762, Adjusted R-squared:  0.0628
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1078 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.205  0.875  0.950  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1          1.017
## Year            1.035 16          1.001

```





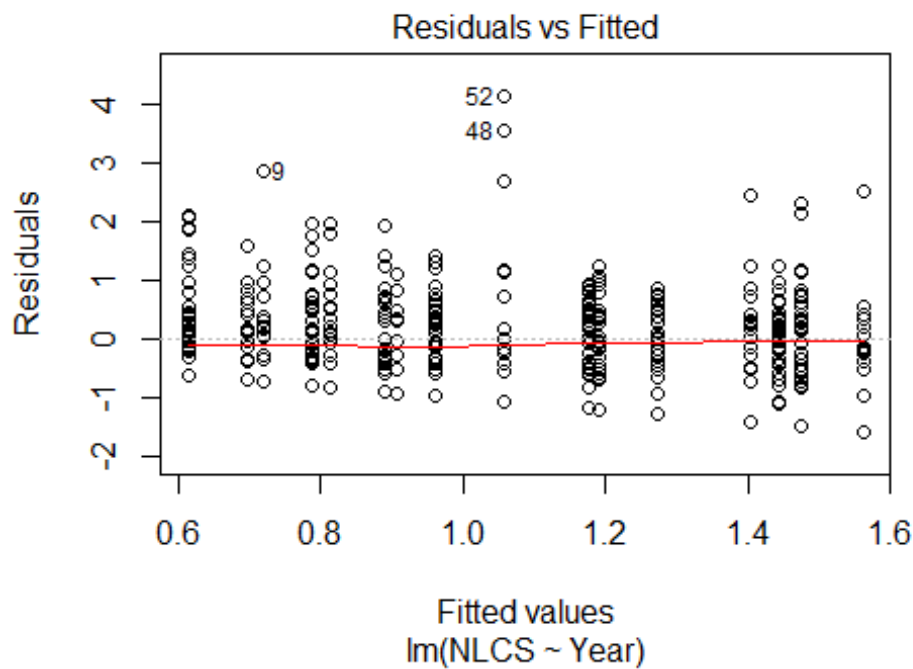
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5273 -0.4051 0.0168 0.4048 2.0297
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4046 0.1008 13.94 < 2e-16 ***
## LastAuthorFemale1 0.0741 0.0464 1.60 0.11012
## Year1997 -0.0240 0.1699 -0.14 0.88751
## Year1998 0.1227 0.1697 0.72 0.46980
## Year1999 -0.4834 0.1416 -3.41 0.00067 ***
## Year2000 -0.2834 0.1336 -2.12 0.03416 *
## Year2001 -0.0104 0.1406 -0.07 0.94121
## Year2002 -0.2698 0.1388 -1.94 0.05220 .
## Year2003 0.0229 0.1381 0.17 0.86822
## Year2004 -0.0778 0.1316 -0.59 0.55473
## Year2005 -0.1683 0.1231 -1.37 0.17169
## Year2006 -0.1691 0.1203 -1.41 0.15992
```

```

## Year2007          -0.2917      0.1169    -2.49  0.01275 *
## Year2008          -0.3837      0.1151    -3.33  0.00088 ***
## Year2009          -0.4365      0.1144    -3.82  0.00014 ***
## Year2010          -0.3056      0.1185    -2.58  0.01002 *
## Year2011          -0.3908      0.1132    -3.45  0.00058 ***
## Year2012          -0.3578      0.1136    -3.15  0.00168 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0685, Adjusted R-squared:  0.055
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1075 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.220  0.872  0.949  0.908  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1187"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2214"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   62   49   49   59   40   37   26   22   39   44   42   55   61   61
## 2011 2012
##   61   65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21   33   36   33   30   21   22   13   14   28   25   28   35   38   37
## 2011 2012

```

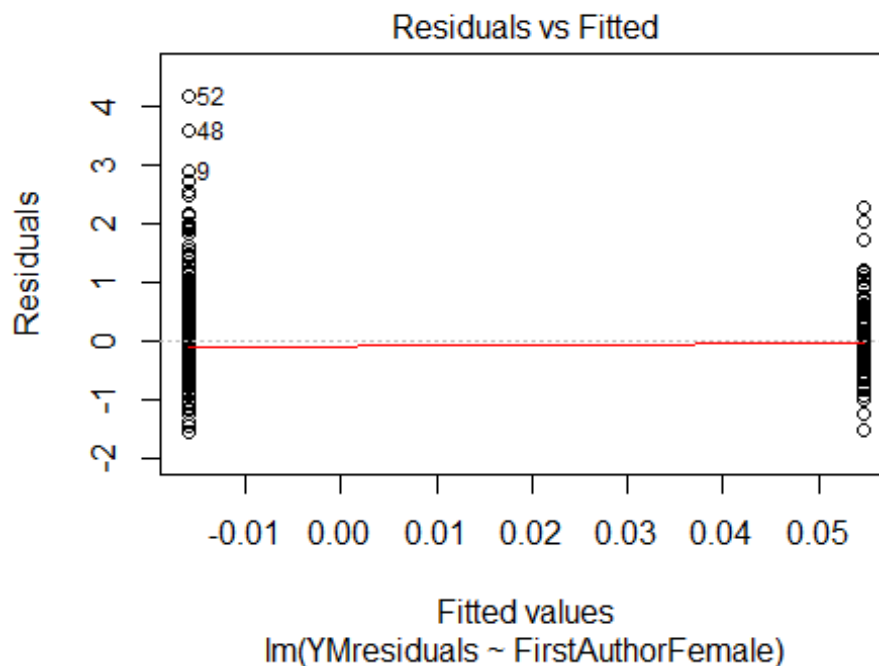
```
## 37 48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 32 35 33 27 19 20 11 12 22 23 27 32 36 29
## 2011 2012
## 33 42
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 60, df = 16, p-value = 6e-07
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 6.3, df = 1, p-value = 0.01
## [1] "Female first author team size 2018 geometric mean: 1.49407039595456"
## [1] "Male first author team size 2018 geometric mean: 2.44822220271585"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.008
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.4186037934323"
## [1] "Male last author team size 2018 geometric mean: 2.49677380096509"

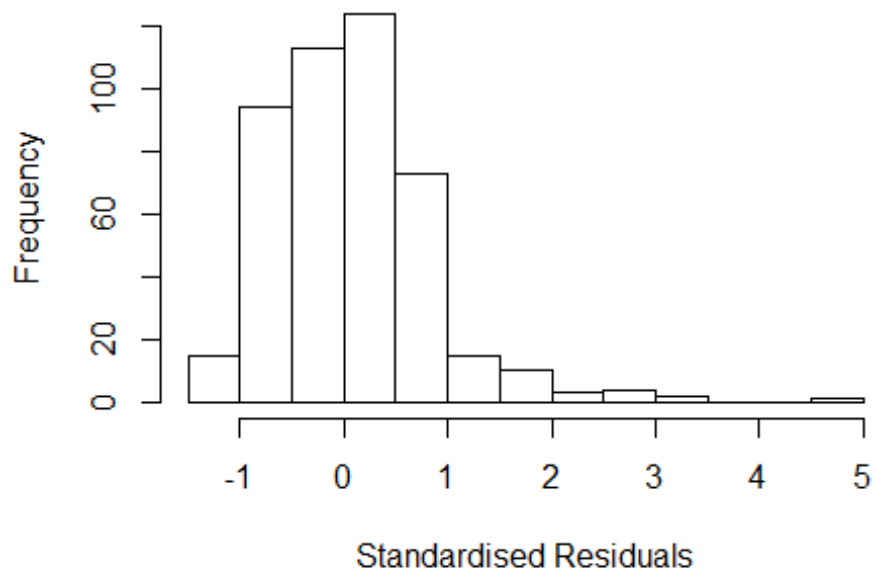
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 110, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	2.360	1	1.536
LastAuthorFemale	2.446	1	1.564
UniqueAuthors	2.009	4	1.091
Year	2.679	16	1.031

## Residuals from first and last author and team size



```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 9      0030262479 3.595 1996      2214      1      3.010
## 50     0031247110 3.740 1997      2214      1      2.883
## 52     0031248512 5.191 1997      2214      1      4.525
## 53     0031250221 3.757 1997      2214      1      3.071
## 305    0035249623 3.850 2001      2214      1      2.537
## 415    0742290132 4.094 2004      2208      2      2.877
## 689    72849107558 3.615 2009      2214      2      2.569
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.48797 -0.48108  0.00492  0.46218  4.52529
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.2646     0.1559   1.70  0.09030 .
## FirstAuthorFemale1 0.0909     0.1078   0.84  0.39952
## LastAuthorFemale1 0.1744     0.1188   1.47  0.14273
## UniqueAuthors2    0.3167     0.0787   4.02 6.8e-05 ***
```

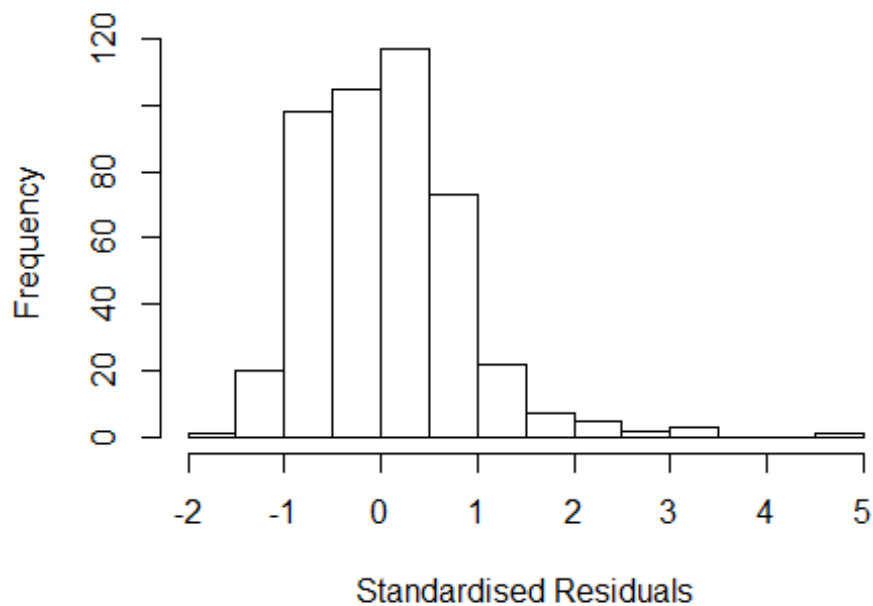
```

## UniqueAuthors3      0.3200      0.1016      3.15  0.00174 **
## UniqueAuthors4      0.3402      0.1619      2.10  0.03616 *
## UniqueAuthors5      0.5111      0.1504      3.40  0.00074 ***
## Year1997             0.0811      0.2008      0.40  0.68651
## Year1998             0.0774      0.1839      0.42  0.67394
## Year1999            -0.0539      0.2026     -0.27  0.79053
## Year2000             0.2417      0.1899      1.27  0.20387
## Year2001             0.7313      0.2489      2.94  0.00348 **
## Year2002             0.0916      0.1899      0.48  0.62975
## Year2003             0.3280      0.2203      1.49  0.13719
## Year2004             0.9521      0.2275      4.19  3.4e-05 ***
## Year2005             0.1181      0.1990      0.59  0.55327
## Year2006             0.7238      0.2071      3.50  0.00052 ***
## Year2007             0.6574      0.2047      3.21  0.00142 **
## Year2008             0.8449      0.1783      4.74  2.9e-06 ***
## Year2009             0.7818      0.2047      3.82  0.00015 ***
## Year2010             0.6414      0.1973      3.25  0.00124 **
## Year2011             0.4294      0.2090      2.05  0.04051 *
## Year2012             0.2165      0.1799      1.20  0.22953
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.656
## Multiple R-squared:  0.269, Adjusted R-squared:  0.232
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(34,35) are outliers with |weight| <= 6.2e-07 ( <
0.00022);
## 36 weights are ~= 1. The remaining 416 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0016 0.8830 0.9460 0.8980 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.20e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

```

##		GVIF	Df	GVIF^(1/(2*Df))
##	FirstAuthorFemale	2.090	1	1.446
##	LastAuthorFemale	2.122	1	1.457
##	Year	1.448	16	1.012

### Residuals from first and last author



```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 9    0030262479 3.595 1996    2214      1      3.109
## 50   0031247110 3.740 1997    2214      1      3.224
## 52   0031248512 5.191 1997    2214      1      4.675
## 53   0031250221 3.757 1997    2214      1      3.241
## 305  0035249623 3.850 2001    2214      1      2.633
## 415  0742290132 4.094 2004    2208      2      2.692
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5184 -0.5158  0.0128  0.4895  4.6752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4862    0.1516   3.21  0.00144 **
## FirstAuthorFemale1 0.0768    0.1042   0.74  0.46165
```

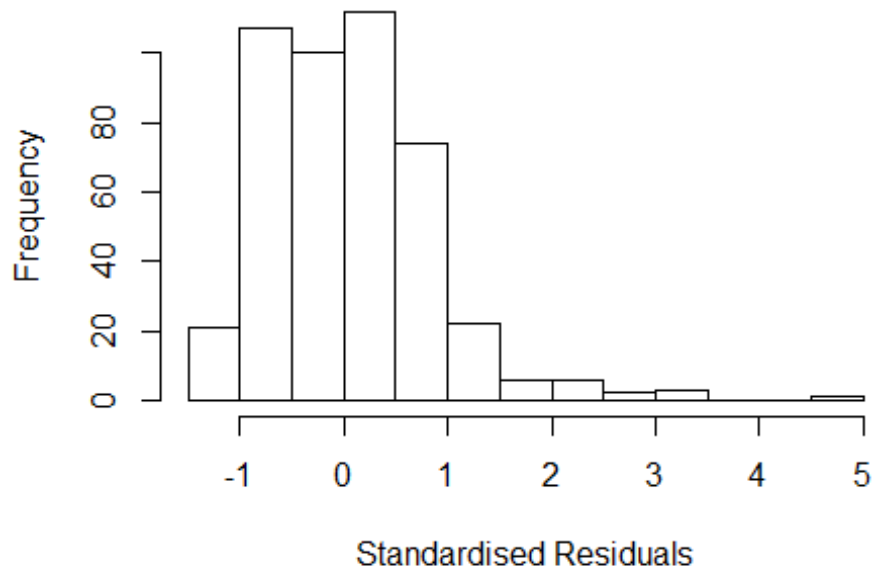
```

## LastAuthorFemale1    0.1175    0.1135    1.04    0.30120
## Year1997             0.0296    0.2007    0.15    0.88283
## Year1998             0.0413    0.1850    0.22    0.82329
## Year1999            -0.0739    0.1941   -0.38    0.70365
## Year2000             0.1689    0.1953    0.86    0.38760
## Year2001             0.7304    0.2590    2.82    0.00503 **
## Year2002             0.0517    0.1959    0.26    0.79180
## Year2003             0.2443    0.2256    1.08    0.27933
## Year2004             0.9162    0.2323    3.94    9.4e-05 ***
## Year2005             0.1195    0.2103    0.57    0.57003
## Year2006             0.7417    0.2029    3.66    0.00029 ***
## Year2007             0.6770    0.2071    3.27    0.00116 **
## Year2008             0.9121    0.1774    5.14    4.1e-07 ***
## Year2009             0.8380    0.2140    3.92    0.00010 ***
## Year2010             0.6464    0.1999    3.23    0.00132 **
## Year2011             0.4286    0.2086    2.05    0.04049 *
## Year2012             0.1190    0.1876    0.63    0.52621
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.225, Adjusted R-squared:  0.193
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(32,34,35) are outliers with |weight| = 0 ( < 0.00022);
## 36 weights are ~= 1. The remaining 415 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.001  0.895   0.946   0.900   0.980   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.20e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.28 1          1.131
## Year              1.28 16          1.008

```



## Residuals from first author



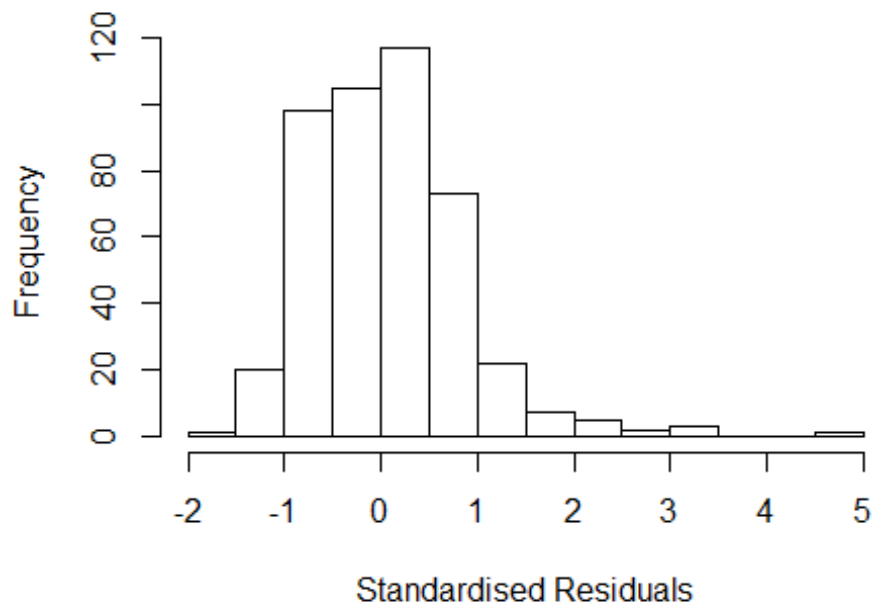
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 9   0030262479 3.595 1996    2214     1    3.109
## 50  0031247110 3.740 1997    2214     1    3.224
## 52  0031248512 5.191 1997    2214     1    4.675
## 53  0031250221 3.757 1997    2214     1    3.241
## 305 0035249623 3.850 2001    2214     1    2.633
## 415 0742290132 4.094 2004    2208     2    2.692
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47710 -0.52148 -0.00588  0.50247  4.66952
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.51875    0.14992   3.46  0.00059 ***
## FirstAuthorFemale1 0.14346    0.08224   1.74  0.08179 .
## Year1997         0.00273    0.19998   0.01  0.98910
## Year1998         0.02010    0.18585   0.11  0.91393
## Year1999        -0.10148    0.19117  -0.53  0.59581
## Year2000         0.14411    0.19545   0.74  0.46133
## Year2001         0.69413    0.25874   2.68  0.00758 **
```

```

## Year2002          0.03068      0.19504      0.16  0.87510
## Year2003          0.21175      0.22410      0.94  0.34523
## Year2004          0.88599      0.23249      3.81  0.00016 ***
## Year2005          0.11179      0.20779      0.54  0.59088
## Year2006          0.71055      0.19876      3.57  0.00039 ***
## Year2007          0.65042      0.20493      3.17  0.00161 **
## Year2008          0.90325      0.17813      5.07  5.9e-07 ***
## Year2009          0.81489      0.21289      3.83  0.00015 ***
## Year2010          0.63290      0.19938      3.17  0.00161 **
## Year2011          0.40061      0.20710      1.93  0.05372 .
## Year2012          0.10755      0.18846      0.57  0.56852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.22, Adjusted R-squared:  0.189
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 3 observations c(32,34,35) are outliers with |weight| = 0 ( < 0.00022);
## 33 weights are ~= 1. The remaining 418 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0062 0.8910 0.9480 0.9030 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.20e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.291 1          1.136
## Year              1.291 16          1.008

```

## Residuals from last author



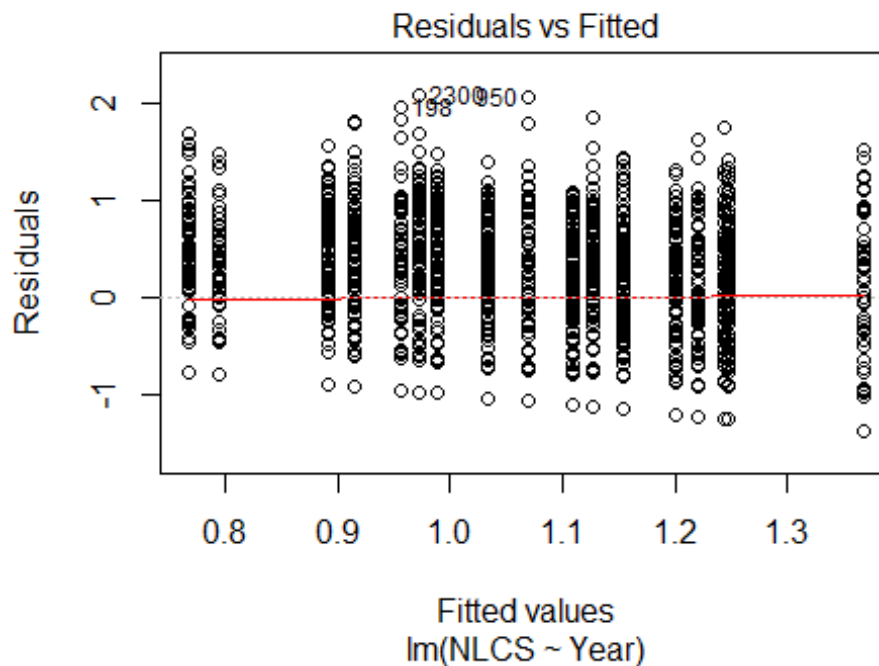
```
## [1] "List of 6 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 9   0030262479 3.595 1996    2214      1    3.109
## 50  0031247110 3.740 1997    2214      1    3.224
## 52  0031248512 5.191 1997    2214      1    4.675
## 53  0031250221 3.757 1997    2214      1    3.241
## 305 0035249623 3.850 2001    2214      1    2.633
## 415 0742290132 4.094 2004    2208      2    2.692
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49939 -0.51761  0.00181  0.49480  4.67339
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4812    0.1522   3.16 0.00167 **
## LastAuthorFemale1 0.1631    0.0892   1.83 0.06818 .
## Year1997        0.0364    0.2014   0.18 0.85659
## Year1998        0.0488    0.1843   0.26 0.79154
## Year1999       -0.0616    0.1935  -0.32 0.75035
## Year2000        0.1776    0.1959   0.91 0.36508
## Year2001        0.7508    0.2575   2.92 0.00373 **
```

```

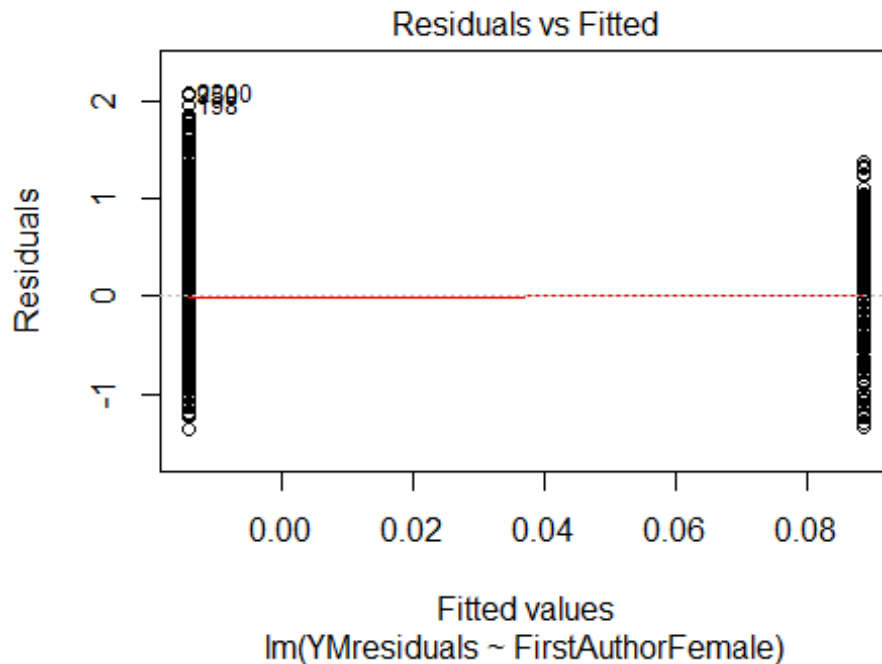
## Year2002          0.0580      0.1975      0.29  0.76910
## Year2003          0.2493      0.2260      1.10  0.27054
## Year2004          0.9211      0.2328      3.96  8.9e-05 ***
## Year2005          0.1324      0.2088      0.63  0.52655
## Year2006          0.7541      0.2016      3.74  0.00021 ***
## Year2007          0.6986      0.2051      3.41  0.00072 ***
## Year2008          0.9216      0.1780      5.18  3.4e-07 ***
## Year2009          0.8551      0.2140      4.00  7.6e-05 ***
## Year2010          0.6662      0.1981      3.36  0.00084 ***
## Year2011          0.4511      0.2041      2.21  0.02757 *
## Year2012          0.1408      0.1859      0.76  0.44929
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.674
## Multiple R-squared:  0.224, Adjusted R-squared:  0.194
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(32,34,35) are outliers with |weight| = 0 ( < 0.00022);
## 39 weights are ~= 1. The remaining 412 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0007 0.8890 0.9460 0.8990 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.20e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 454"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2215"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 282 273 306 282 266 203 188 212 242 257 250 278 255 259 271

```

```
## 2011 2012
## 261 331
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 84 70 92 83 79 54 65 97 127 131 131 147 134 140 139
## 2011 2012
## 138 181
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 68 85 77 75 47 60 86 122 120 115 137 115 121 124
## 2011 2012
## 118 158
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.007
```

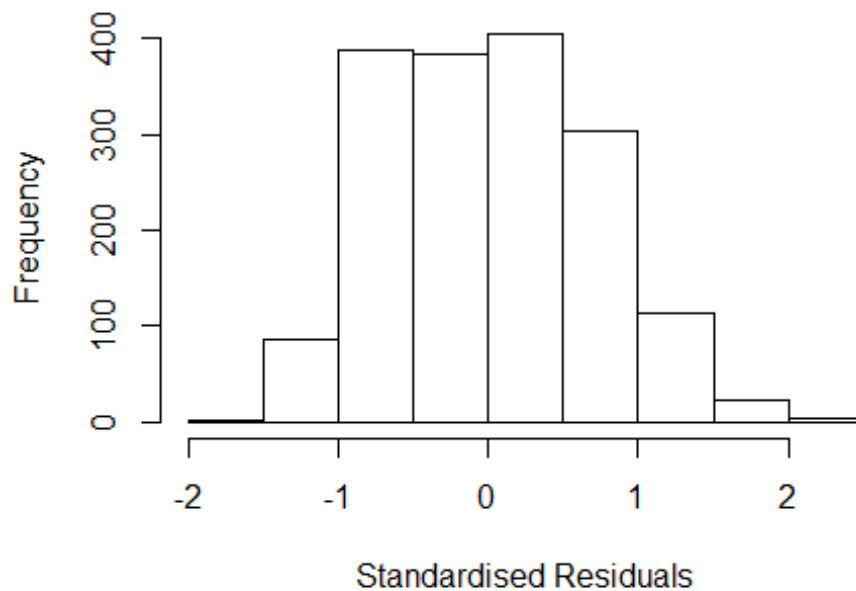


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.2637737160777"
## [1] "Male first author team size 2018 geometric mean: 2.70806440824814"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6600, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.52467863119572"
## [1] "Male last author team size 2018 geometric mean: 2.87634895586255"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.174 1      1.084
## LastAuthorFemale  1.167 1      1.080
## UniqueAuthors    1.260 4      1.029
## Year              1.292 16     1.008
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.590 -0.551 -0.008 0.519 2.415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.71257 0.09865 7.22 7.7e-13 ***
## FirstAuthorFemale1 0.09007 0.05366 1.68 0.093 .
## LastAuthorFemale1 -0.07401 0.05876 -1.26 0.208
## UniqueAuthors2 0.49063 0.04492 10.92 < 2e-16 ***
## UniqueAuthors3 0.56261 0.05792 9.71 < 2e-16 ***
## UniqueAuthors4 0.59477 0.08083 7.36 2.9e-13 ***
## UniqueAuthors5 0.69759 0.08573 8.14 7.8e-16 ***
## Year1997 -0.14282 0.12794 -1.12 0.264
## Year1998 0.16359 0.13234 1.24 0.217
## Year1999 0.06989 0.14099 0.50 0.620
```

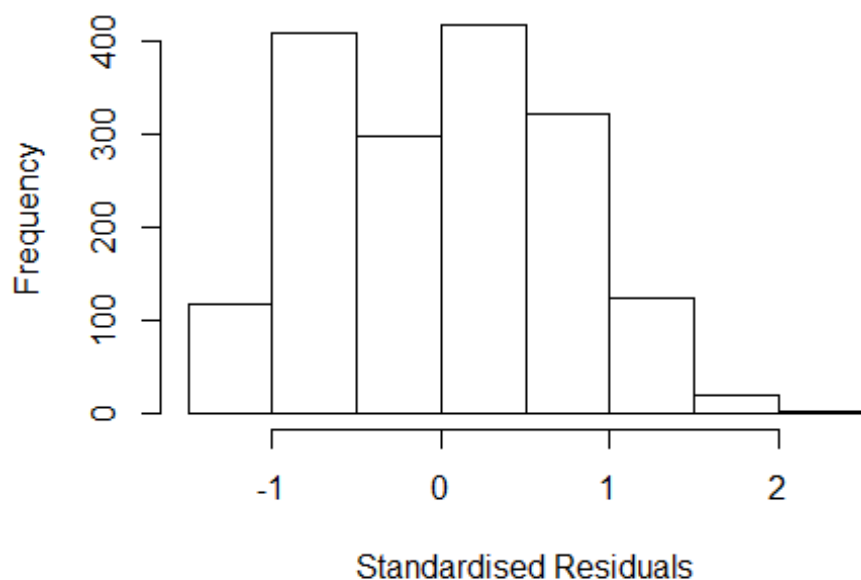
```

## Year2000      0.20603    0.13066    1.58    0.115
## Year2001      0.28280    0.16702    1.69    0.091 .
## Year2002      0.12199    0.14038    0.87    0.385
## Year2003      0.18639    0.12350    1.51    0.131
## Year2004     -0.08505    0.12353   -0.69    0.491
## Year2005     -0.27792    0.11392   -2.44    0.015 *
## Year2006     -0.16181    0.11941   -1.36    0.176
## Year2007     -0.13036    0.11782   -1.11    0.269
## Year2008     -0.02047    0.12325   -0.17    0.868
## Year2009      0.00194    0.11309    0.02    0.986
## Year2010      0.10832    0.11808    0.92    0.359
## Year2011     -0.00419    0.11459   -0.04    0.971
## Year2012     -0.01755    0.11051   -0.16    0.874
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.168, Adjusted R-squared:  0.157
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1589 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.158  0.873  0.934  0.903  0.978  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.178 1      1.085
## LastAuthorFemale  1.158 1      1.076
## Year              1.053 16      1.002

```



## Residuals from first and last author



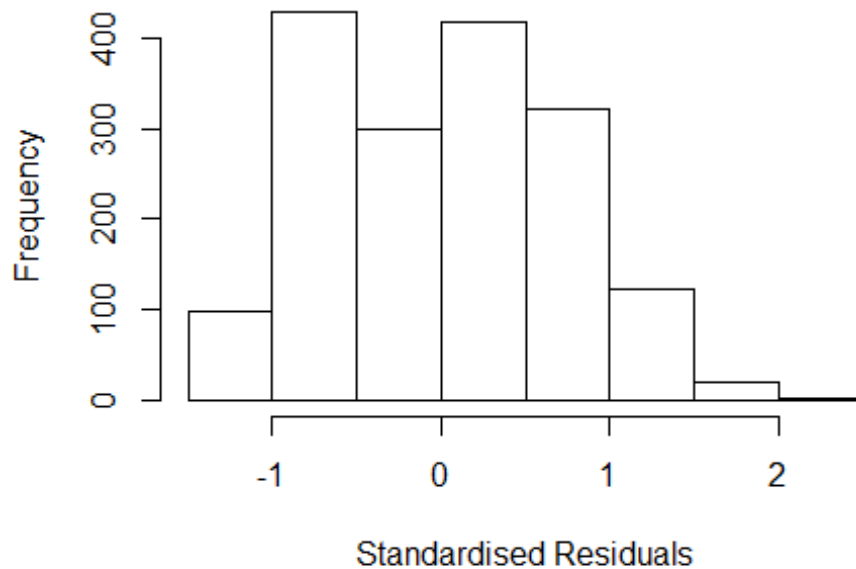
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3344 -0.6779 0.0462 0.5688 2.1344
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9264 0.0961 9.64 <2e-16 ***
## FirstAuthorFemale1 0.1271 0.0551 2.31 0.021 *
## LastAuthorFemale1 -0.0449 0.0586 -0.77 0.443
## Year1997 -0.1515 0.1264 -1.20 0.231
## Year1998 0.1914 0.1311 1.46 0.144
## Year1999 0.0826 0.1374 0.60 0.548
## Year2000 0.2654 0.1261 2.11 0.035 *
## Year2001 0.4080 0.1601 2.55 0.011 *
## Year2002 0.2630 0.1347 1.95 0.051 .
## Year2003 0.2874 0.1227 2.34 0.019 *
## Year2004 -0.0177 0.1295 -0.14 0.891
## Year2005 -0.2404 0.1198 -2.01 0.045 *
```

```

## Year2006          -0.1211      0.1217   -0.99    0.320
## Year2007          -0.0633      0.1210   -0.52    0.601
## Year2008           0.0103      0.1288    0.08    0.936
## Year2009           0.0775      0.1170    0.66    0.508
## Year2010           0.2687      0.1164    2.31    0.021 *
## Year2011           0.1582      0.1139    1.39    0.165
## Year2012           0.1819      0.1101    1.65    0.099 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.809
## Multiple R-squared:  0.0481, Adjusted R-squared:  0.0379
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1590 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.467  0.888  0.941  0.926  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1      1.018
## Year              1.037 16      1.001

```

## Residuals from first author

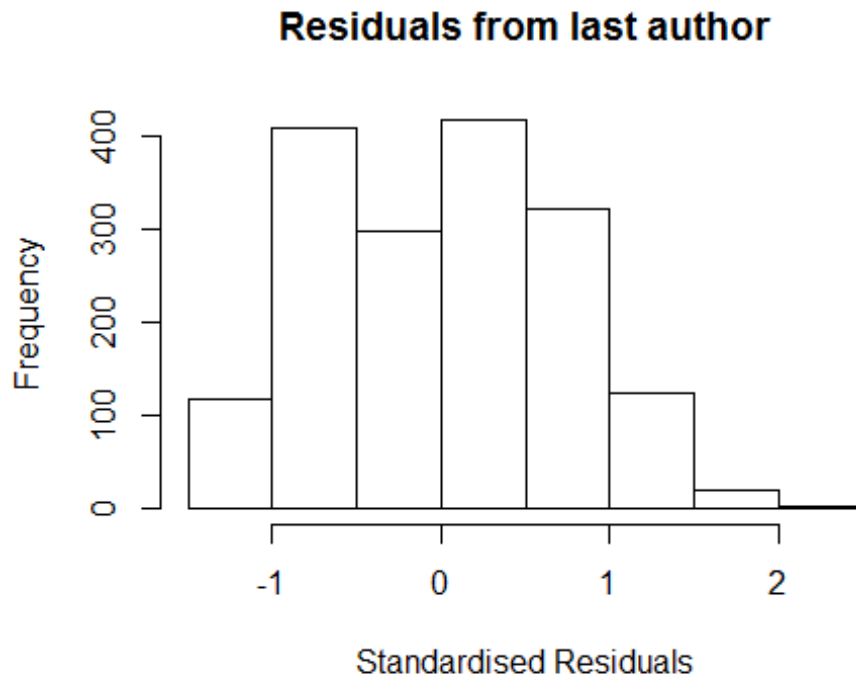


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3327 -0.6805 0.0472 0.5725 2.1381
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92350 0.09665 9.56 <2e-16 ***
## FirstAuthorFemale1 0.10837 0.05243 2.07 0.039 *
## Year1997 -0.15185 0.12687 -1.20 0.232
## Year1998 0.19317 0.13142 1.47 0.142
## Year1999 0.08265 0.13787 0.60 0.549
## Year2000 0.26520 0.12639 2.10 0.036 *
## Year2001 0.40916 0.16073 2.55 0.011 *
## Year2002 0.26241 0.13477 1.95 0.052 .
## Year2003 0.28600 0.12289 2.33 0.020 *
## Year2004 -0.01865 0.12987 -0.14 0.886
## Year2005 -0.24296 0.11993 -2.03 0.043 *
## Year2006 -0.12228 0.12184 -1.00 0.316
```

```

## Year2007      -0.06367    0.12130   -0.52    0.600
## Year2008      0.00968    0.12914    0.07    0.940
## Year2009      0.07526    0.11723    0.64    0.521
## Year2010      0.26663    0.11668    2.29    0.022 *
## Year2011      0.15716    0.11418    1.38    0.169
## Year2012      0.18170    0.11040    1.65    0.100 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.805
## Multiple R-squared:  0.0479, Adjusted R-squared:  0.0383
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1585 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.460  0.888  0.940  0.925  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.011
## Year              1.023 16          1.001

```



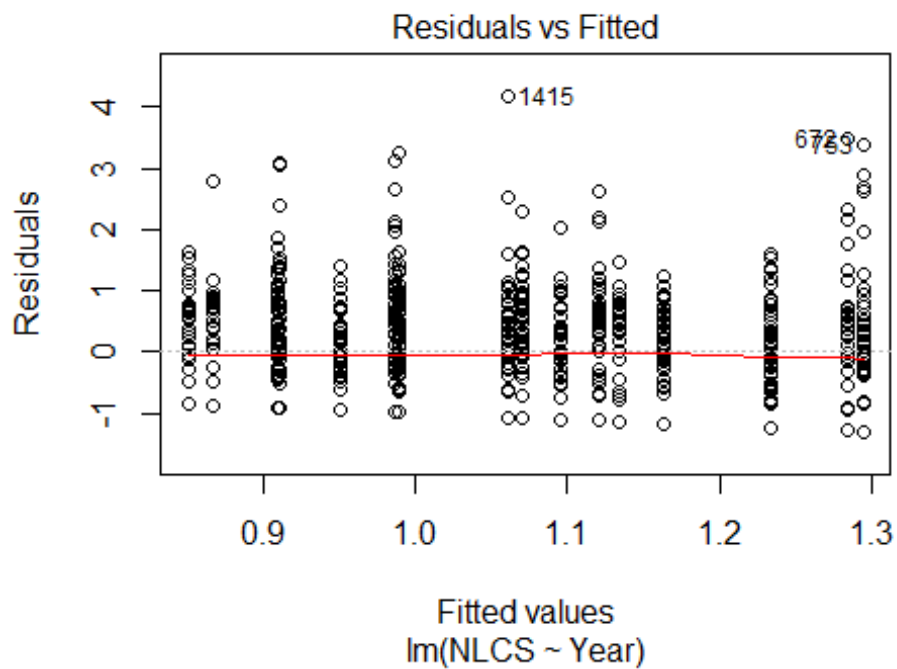
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.334 -0.686 0.039 0.557 2.129
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.92990 0.09736 9.55 <2e-16 ***
## LastAuthorFemale1 0.00912 0.05467 0.17 0.868
## Year1997 -0.15251 0.12778 -1.19 0.233
## Year1998 0.19592 0.13173 1.49 0.137
## Year1999 0.08873 0.13790 0.64 0.520
## Year2000 0.27098 0.12691 2.14 0.033 *
## Year2001 0.40441 0.16168 2.50 0.012 *
## Year2002 0.27240 0.13596 2.00 0.045 *
## Year2003 0.29588 0.12382 2.39 0.017 *
## Year2004 -0.01604 0.13073 -0.12 0.902
## Year2005 -0.22689 0.12074 -1.88 0.060 .
## Year2006 -0.11769 0.12272 -0.96 0.338
```

```

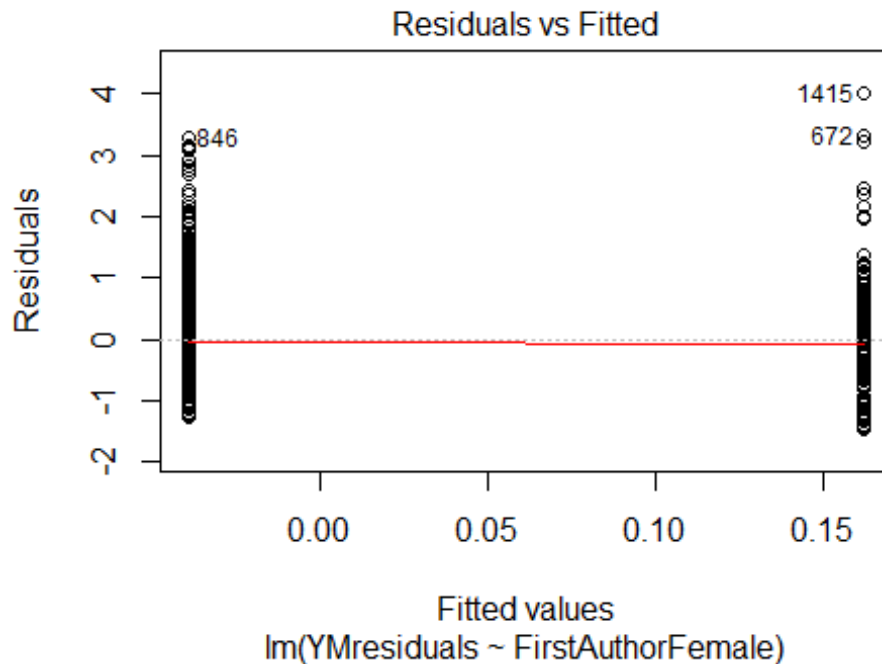
## Year2007          -0.04881      0.12198      -0.40      0.689
## Year2008           0.01240      0.13008       0.10      0.924
## Year2009           0.08011      0.11808       0.68      0.498
## Year2010           0.27534      0.11808       2.33      0.020 *
## Year2011           0.16868      0.11540       1.46      0.144
## Year2012           0.19811      0.11111       1.78      0.075 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.807
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0359
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 126 weights are ~= 1. The remaining 1583 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.466  0.887  0.938  0.925  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1709"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2216"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   80   73   61   67   89   67   69   60   62   81   82  109  127   85   98
## 2011 2012
##   89  101
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   39   28   38   57   40   53   40   42   59   57   77  102   63   70
## 2011 2012

```

```
## 66 76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 38 25 37 53 35 51 37 40 55 56 71 94 59 63
## 2011 2012
## 61 70
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 1e-04
```



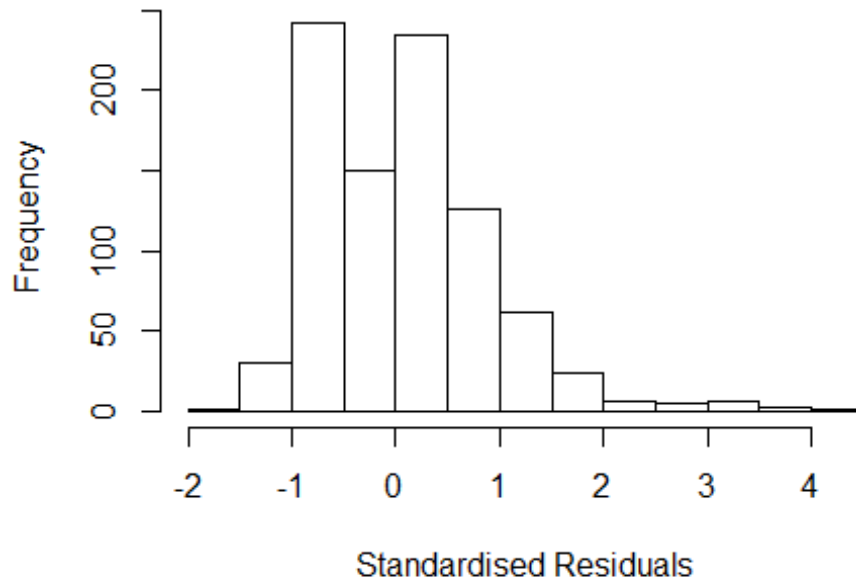
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.35575932868366"
## [1] "Male first author team size 2018 geometric mean: 1.48455702405437"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 970, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.25788403767638"
## [1] "Male last author team size 2018 geometric mean: 1.54726741775754"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 860, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.504 1      1.226
## LastAuthorFemale  1.395 1      1.181
## UniqueAuthors    1.480 4      1.050
## Year              1.518 16     1.013
```



## Residuals from first and last author and team size



```
## [1] "List of 15 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 477  30544433519 3.948 2002    2216      1    3.478
## 522  85007973681 3.313 2002    1213      2    2.843
## 524  85007975083 3.980 2002    1213      2    3.510
## 608  85007986198 3.649 2003    1213      2    3.100
## 672  85008522337 4.779 2004    1213      2    3.819
## 705  61449501605 4.172 2005    1213      2    3.040
## 733  41549159477 3.933 2005    1213      2    3.207
## 735  61449321564 3.262 2005    1213      2    2.536
## 750  85008524788 3.997 2005    1213      2    3.026
## 753  85008541082 4.668 2005    1213      2    3.311
## 846  85008541069 4.250 2006    1213      2    3.564
## 997  65849206437 3.639 2008    1213      2    2.961
## 1009 84973622246 4.091 2008    1206      3    2.916
## 1069 57749130842 3.639 2008    1213      2    2.961
## 1415 79952092664 5.230 2011    2216      1    4.438
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
```

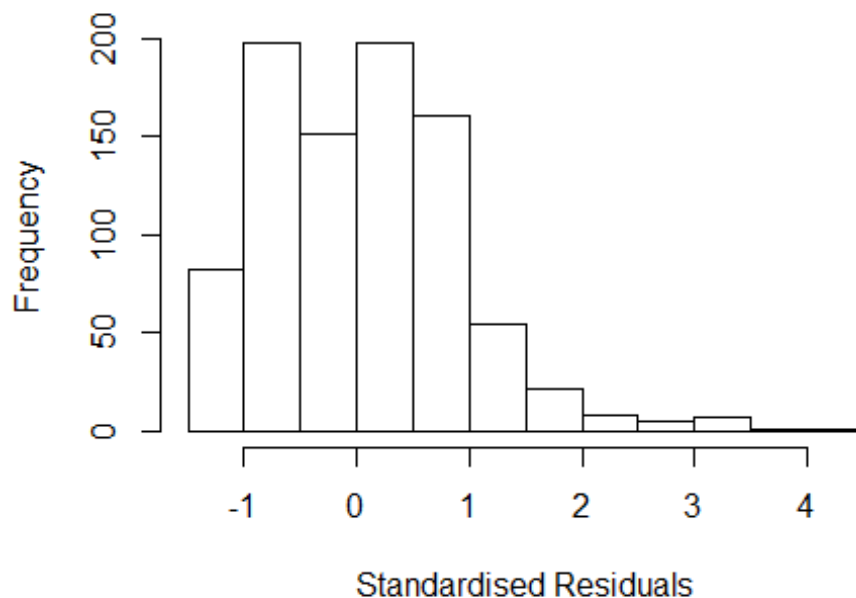
```

## -1.621 -0.650 0.053 0.526 4.438
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.67288    0.12027   5.59 3.0e-08 ***
## FirstAuthorFemale1 0.22517    0.07513   3.00 0.0028 **
## LastAuthorFemale1 -0.16065    0.08330  -1.93 0.0541 .
## UniqueAuthors2    0.40633    0.06587   6.17 1.1e-09 ***
## UniqueAuthors3    0.49673    0.08887   5.59 3.1e-08 ***
## UniqueAuthors4    0.69608    0.11168   6.23 7.1e-10 ***
## UniqueAuthors5    0.89499    0.16586   5.40 8.8e-08 ***
## Year1997          0.05692    0.16894   0.34 0.7362
## Year1998          0.01093    0.20068   0.05 0.9566
## Year1999          0.22994    0.18413   1.25 0.2121
## Year2000          0.24167    0.15009   1.61 0.1077
## Year2001          0.15574    0.17885   0.87 0.3841
## Year2002         -0.20275    0.16436  -1.23 0.2177
## Year2003         -0.12407    0.16974  -0.73 0.4650
## Year2004          0.22260    0.21057   1.06 0.2908
## Year2005          0.05264    0.18110   0.29 0.7714
## Year2006          0.01339    0.15330   0.09 0.9304
## Year2007          0.22931    0.16002   1.43 0.1522
## Year2008          0.00552    0.15179   0.04 0.9710
## Year2009          0.23009    0.14706   1.56 0.1180
## Year2010          0.05913    0.15539   0.38 0.7037
## Year2011          0.05433    0.15146   0.36 0.7199
## Year2012         -0.09690    0.14624  -0.66 0.5078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.759
## Multiple R-squared: 0.139, Adjusted R-squared: 0.117
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 3 observations c(346,458,802) are outliers with |weight| = 0 ( <
0.00011);
## 49 weights are ~= 1. The remaining 838 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0006 0.8770 0.9370 0.9020 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.12e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd

```

```
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.536 1          1.239
## LastAuthorFemale 1.499 1          1.225
## Year              1.155 16          1.005
```

### Residuals from first and last author



```
## [1] "List of 14 outliers with residuals above 2.5"
##          ScopusId NLCS Year OneField Fields residuals
## 477 30544433519 3.948 2002 2216 1 3.269
## 522 85007973681 3.313 2002 1213 2 2.634
## 524 85007975083 3.980 2002 1213 2 3.301
## 608 85007986198 3.649 2003 1213 2 2.932
## 672 85008522337 4.779 2004 1213 2 3.561
## 705 61449501605 4.172 2005 1213 2 3.127
## 733 41549159477 3.933 2005 1213 2 2.888
## 750 85008524788 3.997 2005 1213 2 3.152
## 753 85008541082 4.668 2005 1213 2 3.321
## 846 85008541069 4.250 2006 1213 2 3.352
## 997 65849206437 3.639 2008 1213 2 2.789
## 1009 84973622246 4.091 2008 1206 3 3.241
```

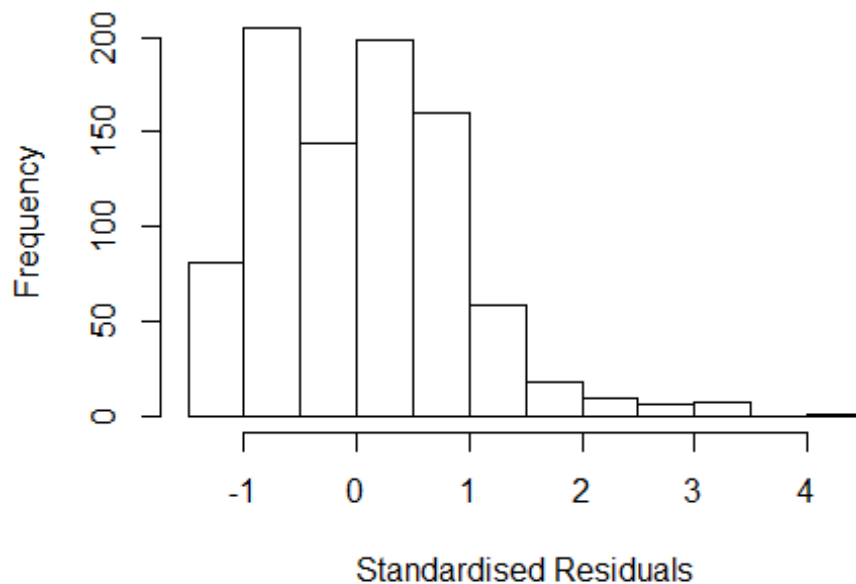
```

## 1069 57749130842 3.639 2008      1213      2      2.789
## 1415 79952092664 5.230 2011      2216      1      4.205
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.272 -0.724  0.023  0.600  4.205
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8134    0.1270   6.40 2.5e-10 ***
## FirstAuthorFemale1 0.3018    0.0827   3.65 0.00028 ***
## LastAuthorFemale1 -0.1998    0.0918  -2.18 0.02981 *
## Year1997         0.0896    0.1646   0.54 0.58629
## Year1998         0.1068    0.2008   0.53 0.59510
## Year1999         0.2561    0.1851   1.38 0.16697
## Year2000         0.3200    0.1592   2.01 0.04478 *
## Year2001         0.1975    0.1926   1.03 0.30547
## Year2002        -0.1347    0.1864  -0.72 0.46994
## Year2003        -0.0968    0.1936  -0.50 0.61722
## Year2004         0.3022    0.2239   1.35 0.17755
## Year2005         0.2318    0.1748   1.33 0.18519
## Year2006         0.0851    0.1652   0.52 0.60655
## Year2007         0.2361    0.1719   1.37 0.16998
## Year2008         0.0366    0.1643   0.22 0.82391
## Year2009         0.3568    0.1601   2.23 0.02610 *
## Year2010         0.1421    0.1706   0.83 0.40515
## Year2011         0.1099    0.1626   0.68 0.49944
## Year2012        -0.0421    0.1554  -0.27 0.78681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.868
## Multiple R-squared:  0.0406, Adjusted R-squared:  0.0208
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 802 is an outlier with |weight| = 0 ( < 0.00011);
## 68 weights are ~ 1. The remaining 821 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0544 0.8930 0.9390 0.9130 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.12e-04      1.82e-12

```

```
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
##  nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##          500          50          2          1          1000          200
##  trace.lev      mts    compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1          1.045
## Year              1.091 16          1.003
```

### Residuals from first author



```
## [1] "List of 14 outliers with residuals above 2.5"
##          ScopusId  NLCS Year OneField Fields residuals
## 477 30544433519 3.948 2002    2216     1    3.269
## 522 85007973681 3.313 2002    1213     2    2.634
## 524 85007975083 3.980 2002    1213     2    3.301
## 608 85007986198 3.649 2003    1213     2    2.932
## 672 85008522337 4.779 2004    1213     2    3.561
## 705 61449501605 4.172 2005    1213     2    3.127
## 733 41549159477 3.933 2005    1213     2    2.888
## 750 85008524788 3.997 2005    1213     2    3.152
## 753 85008541082 4.668 2005    1213     2    3.321
```

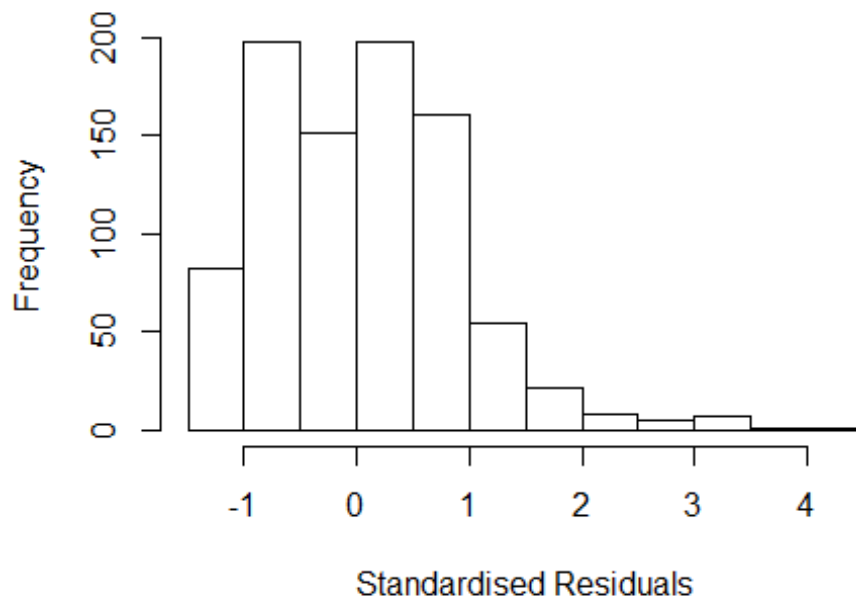
```

## 846 85008541069 4.250 2006 1213 2 3.352
## 997 65849206437 3.639 2008 1213 2 2.789
## 1009 84973622246 4.091 2008 1206 3 3.241
## 1069 57749130842 3.639 2008 1213 2 2.789
## 1415 79952092664 5.230 2011 2216 1 4.205
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3374 -0.7502  0.0364  0.6099  4.1270
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8050    0.1274   6.32 4.2e-10 ***
## FirstAuthorFemale1 0.1839    0.0727   2.53  0.012 *
## Year1997        0.0857    0.1664   0.51  0.607
## Year1998        0.1040    0.2013   0.52  0.605
## Year1999        0.2514    0.1850   1.36  0.174
## Year2000        0.3307    0.1605   2.06  0.040 *
## Year2001        0.1942    0.1929   1.01  0.314
## Year2002       -0.1314    0.1872  -0.70  0.483
## Year2003       -0.1053    0.1936  -0.54  0.587
## Year2004        0.2915    0.2232   1.31  0.192
## Year2005        0.2133    0.1763   1.21  0.227
## Year2006        0.0735    0.1654   0.44  0.657
## Year2007        0.2213    0.1717   1.29  0.198
## Year2008        0.0347    0.1651   0.21  0.834
## Year2009        0.3485    0.1603   2.17  0.030 *
## Year2010        0.1469    0.1707   0.86  0.390
## Year2011        0.1141    0.1632   0.70  0.485
## Year2012       -0.0548    0.1552  -0.35  0.724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.868
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0168
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 802 is an outlier with |weight| = 0 ( < 0.00011);
## 66 weights are ~= 1. The remaining 823 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0678 0.8930 0.9400 0.9120 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00        5.00e-01      4.69e+00      1.00e-07

```

```
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          1.12e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##          500         50         2         1         1000         200
## trace.lev    mts    compute.rd
##           0      1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.06 1          1.030
## Year              1.06 16          1.002
```

### Residuals from last author



```
## [1] "List of 14 outliers with residuals above 2.5"
##           ScopusId NLCS Year OneField Fields residuals
## 477 30544433519 3.948 2002 2216 1 3.269
## 522 85007973681 3.313 2002 1213 2 2.634
## 524 85007975083 3.980 2002 1213 2 3.301
## 608 85007986198 3.649 2003 1213 2 2.932
## 672 85008522337 4.779 2004 1213 2 3.561
## 705 61449501605 4.172 2005 1213 2 3.127
## 733 41549159477 3.933 2005 1213 2 2.888
```

```

## 750 85008524788 3.997 2005 1213 2 3.152
## 753 85008541082 4.668 2005 1213 2 3.321
## 846 85008541069 4.250 2006 1213 2 3.352
## 997 65849206437 3.639 2008 1213 2 2.789
## 1009 84973622246 4.091 2008 1206 3 3.241
## 1069 57749130842 3.639 2008 1213 2 2.789
## 1415 79952092664 5.230 2011 2216 1 4.205
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1994 -0.7438  0.0377  0.6124  4.2666
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.82676    0.12679   6.52 1.2e-10 ***
## LastAuthorFemale1 -0.00971    0.07534  -0.13  0.897
## Year1997        0.08068    0.16628   0.49  0.628
## Year1998        0.10645    0.20184   0.53  0.598
## Year1999        0.26510    0.18520   1.43  0.153
## Year2000        0.34946    0.15981   2.19  0.029 *
## Year2001        0.19894    0.19018   1.05  0.296
## Year2002       -0.11247    0.18659  -0.60  0.547
## Year2003       -0.08296    0.19396  -0.43  0.669
## Year2004        0.29320    0.22430   1.31  0.192
## Year2005        0.23445    0.17481   1.34  0.180
## Year2006        0.08841    0.16474   0.54  0.592
## Year2007        0.25253    0.17219   1.47  0.143
## Year2008        0.03199    0.16430   0.19  0.846
## Year2009        0.37268    0.15940   2.34  0.020 *
## Year2010        0.17425    0.17162   1.02  0.310
## Year2011        0.14636    0.16359   0.89  0.371
## Year2012       -0.01016    0.15446  -0.07  0.948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.875
## Multiple R-squared:  0.0276, Adjusted R-squared:  0.0086
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 802 is an outlier with |weight| = 0 ( < 0.00011);
## 61 weights are ~= 1. The remaining 828 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0397 0.8910 0.9380 0.9140 0.9830 0.9990
## Algorithmic parameters:

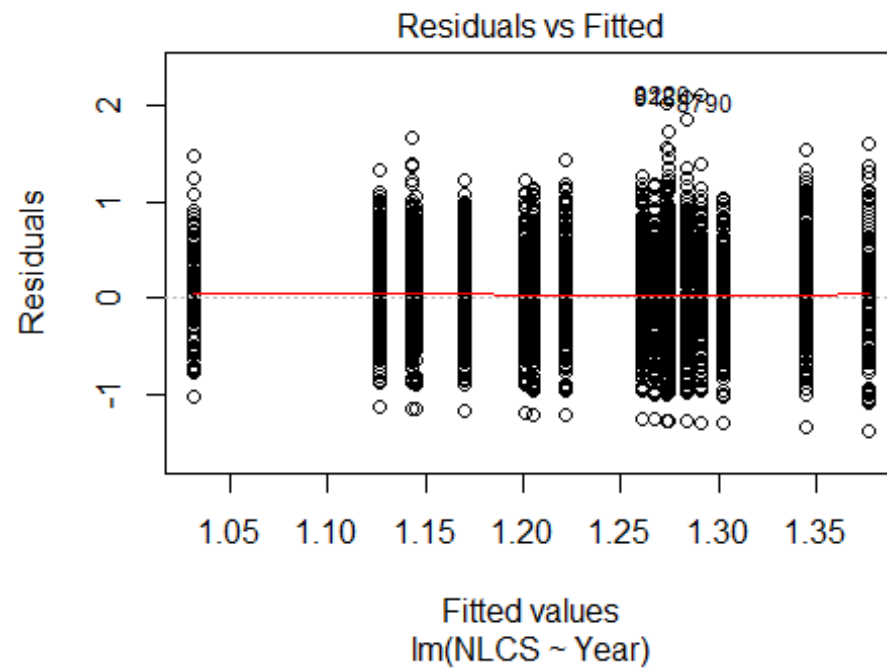
```



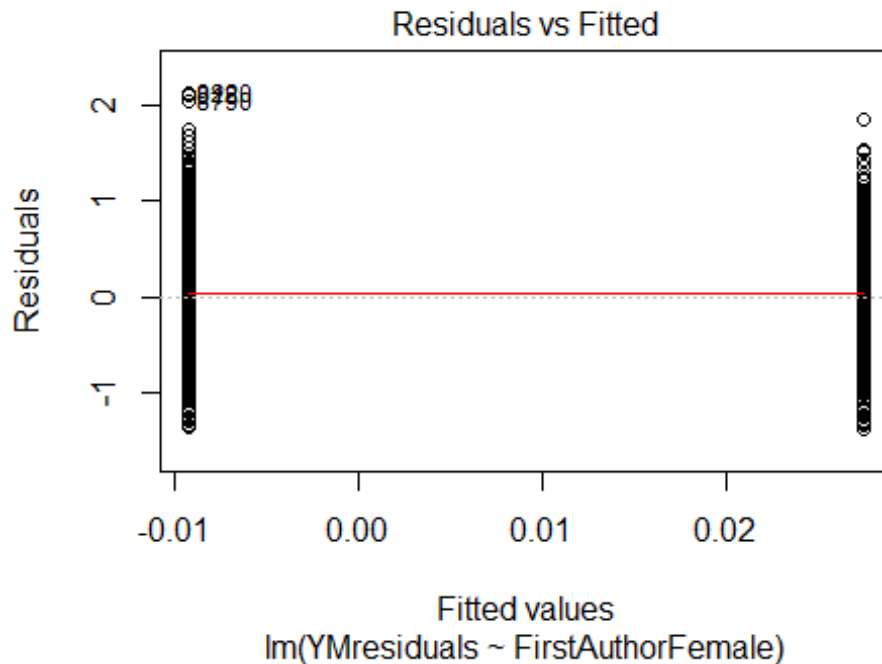
```

##          tuning.chi          bb          tuning.psi          refine.tol
##          1.55e+00          5.00e-01          4.69e+00          1.00e-07
##          rel.tol          solve.tol          eps.outlier          eps.x
##          1.00e-07          1.00e-07          1.12e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##          5.00e-01          5.00e-01
## nResample          max.it          best.r.s          k.fast.s          k.max          maxit.scale
##          500          50          2          1          1000          200
## trace.lev          mts          compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##          "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 890"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2300"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 470 573 483 476 586 544 517 462 424 407 551 523 445 397 377
## 2011 2012
## 385 469
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 151 204 182 254 213 136 332 319 285 282 356 349 274 250 244
## 2011 2012
## 256 318
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 137 192 169 240 199 130 304 302 269 264 317 306 250 230 221
## 2011 2012
## 236 294
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 2e-06

```

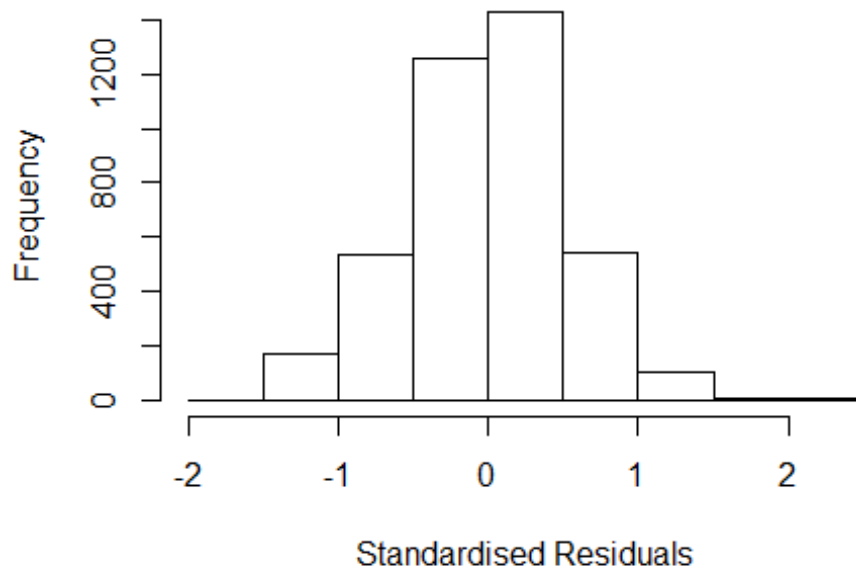


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 8e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 2.57335982921239"
## [1] "Male first author team size 2018 geometric mean: 2.55694830956016"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.172533489423"
## [1] "Male last author team size 2018 geometric mean: 2.70133864043911"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6700, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.048 1      1.024
## LastAuthorFemale  1.036 1      1.018
## UniqueAuthors     1.143 4      1.017
## Year              1.184 16     1.005
```

## Residuals from first and last author and team size



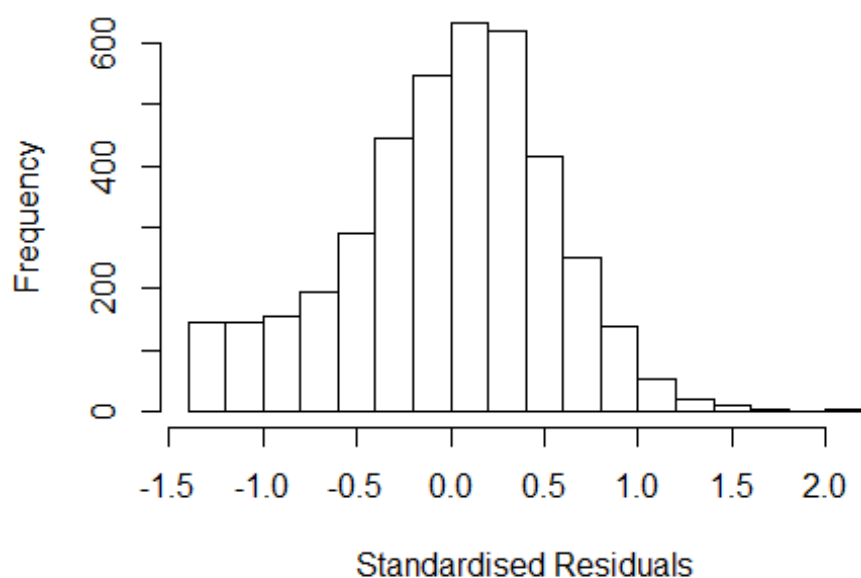
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5376 -0.3365 0.0181 0.3402 2.2784
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22e+00 6.53e-02 18.63 < 2e-16 ***
## FirstAuthorFemale1 -1.32e-02 1.90e-02 -0.70 0.48650
## LastAuthorFemale1 6.79e-05 2.11e-02 0.00 0.99743
## UniqueAuthors2 3.13e-01 2.43e-02 12.89 < 2e-16 ***
## UniqueAuthors3 3.61e-01 2.64e-02 13.70 < 2e-16 ***
## UniqueAuthors4 4.25e-01 2.98e-02 14.27 < 2e-16 ***
## UniqueAuthors5 4.32e-01 2.74e-02 15.78 < 2e-16 ***
## Year1997 -1.11e-01 7.84e-02 -1.42 0.15645
## Year1998 -2.06e-01 7.82e-02 -2.63 0.00858 **
## Year1999 -1.67e-01 7.14e-02 -2.33 0.01969 *
```

```

## Year2000      -1.41e-01  7.18e-02  -1.96  0.04963 *
## Year2001      -3.94e-01  7.99e-02  -4.93  8.7e-07 ***
## Year2002      -2.56e-01  7.19e-02  -3.56  0.00038 ***
## Year2003      -2.93e-01  7.03e-02  -4.17  3.2e-05 ***
## Year2004      -2.69e-01  7.10e-02  -3.79  0.00015 ***
## Year2005      -2.98e-01  7.17e-02  -4.16  3.3e-05 ***
## Year2006      -2.38e-01  6.89e-02  -3.46  0.00056 ***
## Year2007      -2.54e-01  7.00e-02  -3.63  0.00029 ***
## Year2008      -2.04e-01  7.13e-02  -2.87  0.00413 **
## Year2009      -1.48e-01  7.41e-02  -2.00  0.04609 *
## Year2010      -2.00e-01  7.48e-02  -2.68  0.00746 **
## Year2011      -1.66e-01  7.50e-02  -2.21  0.02683 *
## Year2012      -1.22e-01  7.23e-02  -1.68  0.09247 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.114, Adjusted R-squared:  0.109
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 365 weights are ~= 1. The remaining 3695 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0056 0.8530 0.9480 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.46e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.049 1 1.024
## LastAuthorFemale 1.026 1 1.013
## Year 1.045 16 1.001

```

## Residuals from first and last author



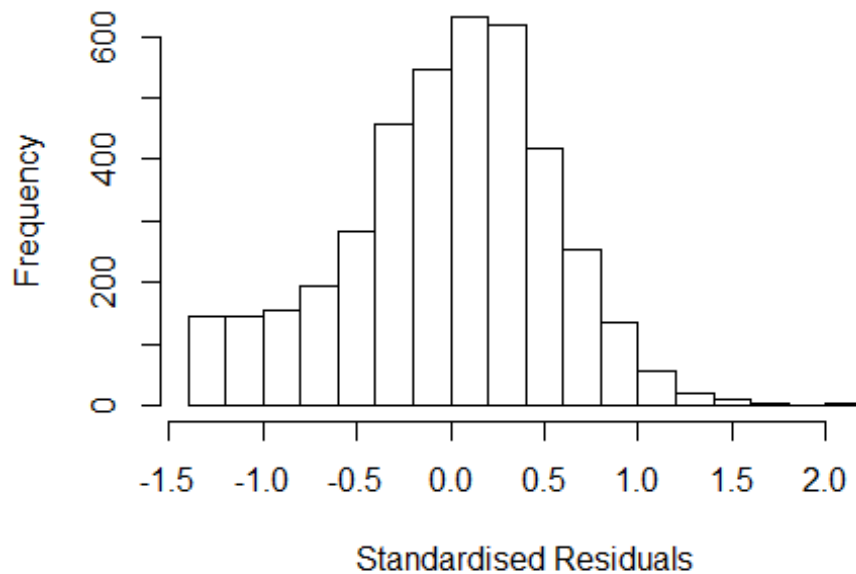
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3891 -0.3484 0.0315 0.3572 2.0960
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3882 0.0649 21.39 < 2e-16 ***
## FirstAuthorFemale1 0.0241 0.0198 1.22 0.2243
## LastAuthorFemale1 -0.0199 0.0221 -0.90 0.3676
## Year1997 -0.1202 0.0796 -1.51 0.1310
## Year1998 -0.1886 0.0783 -2.41 0.0161 *
## Year1999 -0.0974 0.0723 -1.35 0.1780
## Year2000 -0.0656 0.0734 -0.89 0.3717
## Year2001 -0.3478 0.0827 -4.20 2.7e-05 ***
## Year2002 -0.2243 0.0749 -2.99 0.0028 **
## Year2003 -0.2088 0.0719 -2.90 0.0037 **
## Year2004 -0.1958 0.0729 -2.69 0.0073 **
## Year2005 -0.2294 0.0744 -3.08 0.0021 **
```

```

## Year2006          -0.1511      0.0705   -2.14   0.0323 *
## Year2007          -0.1450      0.0706   -2.05   0.0401 *
## Year2008          -0.1058      0.0728   -1.45   0.1463
## Year2009          -0.0666      0.0752   -0.89   0.3757
## Year2010          -0.1164      0.0771   -1.51   0.1311
## Year2011          -0.0886      0.0766   -1.16   0.2476
## Year2012          -0.0232      0.0728   -0.32   0.7498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.0142
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 356 weights are ~= 1. The remaining 3704 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0706 0.8540 0.9500 0.8930 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1      1.016
## Year              1.033 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3852 -0.3511  0.0319  0.3566  2.0985
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3852     0.0646   21.45 < 2e-16 ***
## FirstAuthorFemale1  0.0189     0.0200    0.94  0.3453
## Year1997         -0.1198     0.0795   -1.51  0.1321
## Year1998         -0.1879     0.0782   -2.40  0.0163 *
## Year1999         -0.0972     0.0723   -1.34  0.1787
## Year2000         -0.0654     0.0734   -0.89  0.3730
## Year2001         -0.3472     0.0828   -4.19 2.8e-05 ***
## Year2002         -0.2230     0.0748   -2.98  0.0029 **
## Year2003         -0.2083     0.0719   -2.90  0.0038 **
## Year2004         -0.1952     0.0729   -2.68  0.0074 **
## Year2005         -0.2282     0.0743   -3.07  0.0022 **
## Year2006         -0.1509     0.0705   -2.14  0.0323 *
```

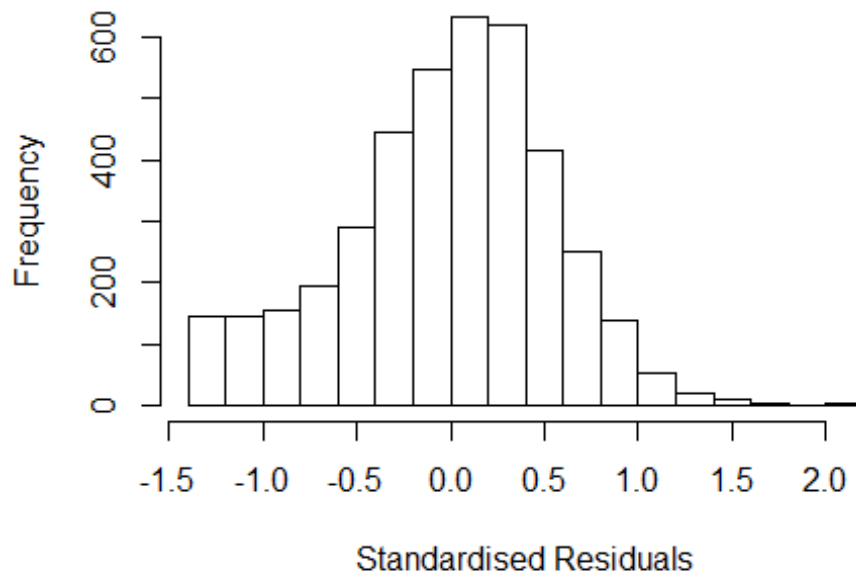


```

## Year2007          -0.1452      0.0706   -2.06   0.0398 *
## Year2008          -0.1053      0.0728   -1.45   0.1479
## Year2009          -0.0661      0.0751   -0.88   0.3788
## Year2010          -0.1158      0.0770   -1.50   0.1328
## Year2011          -0.0893      0.0766   -1.17   0.2438
## Year2012          -0.0239      0.0728   -0.33   0.7428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.0142
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 358 weights are ~= 1. The remaining 3702 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0697 0.8530 0.9490 0.8930 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year            1.012 16          1.000

```

## Residuals from last author



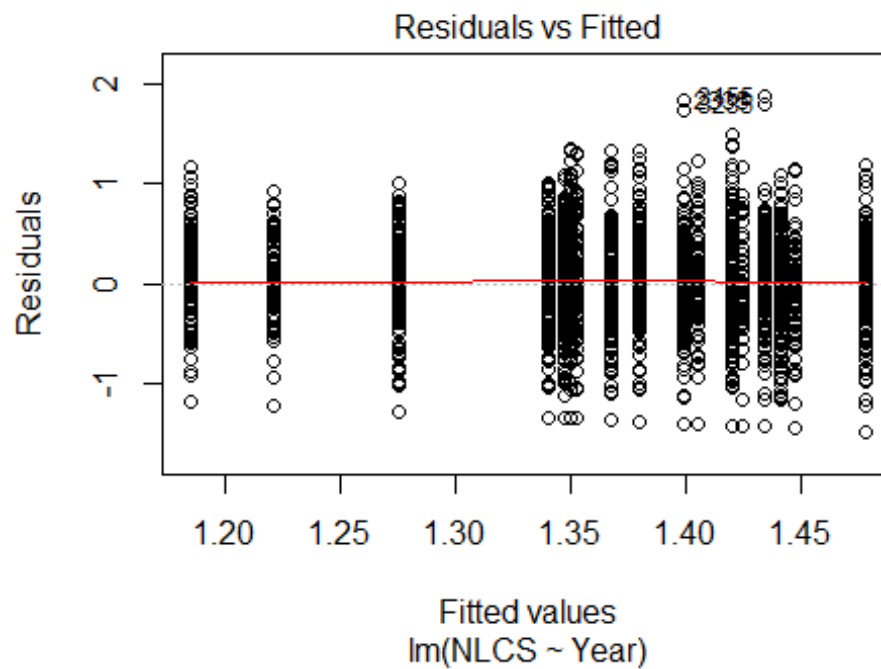
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3912 -0.3473  0.0315  0.3557  2.0924
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3912     0.0647   21.49  < 2e-16 ***
## LastAuthorFemale1 -0.0124     0.0221   -0.56  0.5766
## Year1997         -0.1197     0.0795   -1.51  0.1323
## Year1998         -0.1869     0.0782   -2.39  0.0169 *
## Year1999         -0.0968     0.0722   -1.34  0.1805
## Year2000         -0.0655     0.0734   -0.89  0.3721
## Year2001         -0.3475     0.0827   -4.20  2.7e-05 ***
## Year2002         -0.2231     0.0749   -2.98  0.0029 **
## Year2003         -0.2071     0.0718   -2.88  0.0040 **
## Year2004         -0.1945     0.0729   -2.67  0.0076 **
## Year2005         -0.2275     0.0744   -3.06  0.0022 **
## Year2006         -0.1485     0.0704   -2.11  0.0350 *
```

```

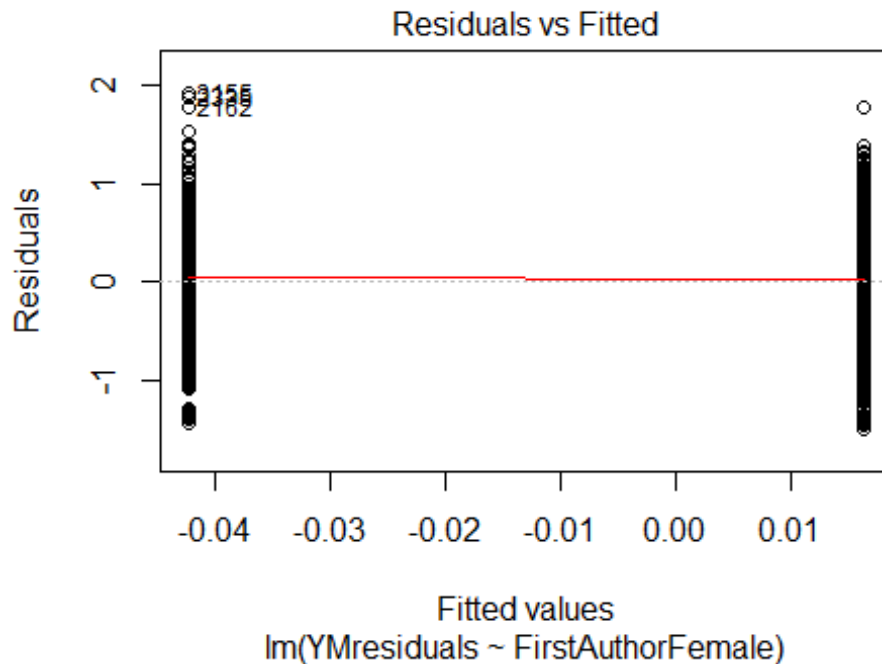
## Year2007          -0.1418      0.0705    -2.01    0.0443 *
## Year2008          -0.1029      0.0727    -1.41    0.1574
## Year2009          -0.0639      0.0750    -0.85    0.3943
## Year2010          -0.1144      0.0771    -1.48    0.1380
## Year2011          -0.0872      0.0766    -1.14    0.2547
## Year2012          -0.0205      0.0727    -0.28    0.7779
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.0183, Adjusted R-squared:  0.0141
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 347 weights are ~= 1. The remaining 3713 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0722 0.8550 0.9500 0.8930 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4060"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2301"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 164 157 191 231 209 233 183 244 208 173 200 209 192 217 215
## 2011 2012
## 238 277
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 101 84 83 128 100 52 149 216 183 156 182 187 168 196 187
## 2011 2012

```

```
## 215 249
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 100 83 80 124 95 45 141 207 173 150 174 179 164 191 175
## 2011 2012
## 204 239
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 52, df = 16, p-value = 1e-05
```

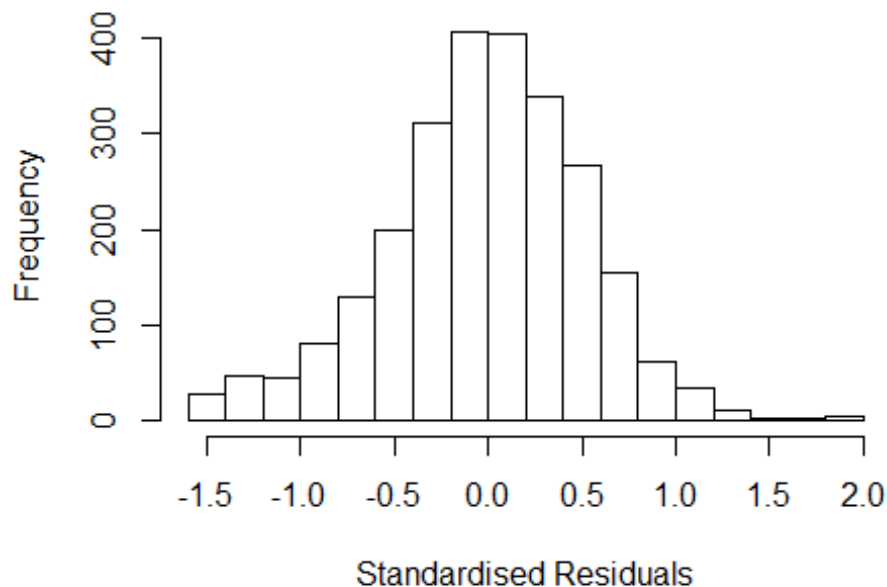


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.3, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 1.49921309263255"
## [1] "Male first author team size 2018 geometric mean: 1.90184835672714"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.55858693228022"
## [1] "Male last author team size 2018 geometric mean: 1.88645199154946"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.462 1          1.209
## LastAuthorFemale  1.500 1          1.225
## UniqueAuthors     1.122 4          1.015
## Year               1.237 16         1.007
```

## Residuals from first and last author and team size



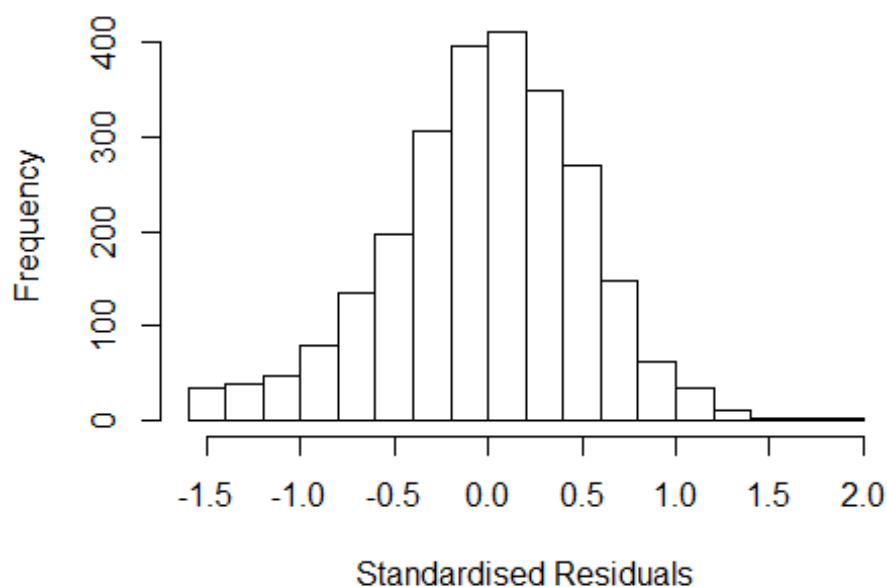
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.49375 -0.31197 0.00786 0.32803 1.90851
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36817 0.06630 20.64 <2e-16 ***
## FirstAuthorFemale1 -0.03863 0.02817 -1.37 0.170
## LastAuthorFemale1 -0.02696 0.02846 -0.95 0.343
## UniqueAuthors2 0.04483 0.02362 1.90 0.058 .
## UniqueAuthors3 0.02839 0.03249 0.87 0.382
## UniqueAuthors4 0.00528 0.06461 0.08 0.935
## UniqueAuthors5 0.08613 0.07148 1.20 0.228
## Year1997 0.04368 0.08504 0.51 0.608
## Year1998 -0.11022 0.08416 -1.31 0.190
## Year1999 -0.15960 0.08292 -1.92 0.054 .
```

```

## Year2000      0.10253    0.08240    1.24    0.214
## Year2001      0.02910    0.09721    0.30    0.765
## Year2002     -0.07801    0.07645   -1.02    0.308
## Year2003     -0.01225    0.07493   -0.16    0.870
## Year2004      0.00305    0.07514    0.04    0.968
## Year2005      0.02754    0.08033    0.34    0.732
## Year2006      0.01326    0.07479    0.18    0.859
## Year2007      0.09181    0.07505    1.22    0.221
## Year2008      0.12559    0.07790    1.61    0.107
## Year2009      0.06573    0.07829    0.84    0.401
## Year2010      0.09892    0.07572    1.31    0.192
## Year2011      0.02418    0.07954    0.30    0.761
## Year2012      0.08287    0.07688    1.08    0.281
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0236, Adjusted R-squared:  0.015
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 2313 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0919 0.8650 0.9500 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.96e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.470 1 1.212
## LastAuthorFemale 1.501 1 1.225
## Year 1.112 16 1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5086 -0.3183 0.0146 0.3304 1.8928
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37702 0.06625 20.78 <2e-16 ***
## FirstAuthorFemale1 -0.03840 0.02830 -1.36 0.175
## LastAuthorFemale1 -0.02670 0.02854 -0.94 0.350
## Year1997 0.04783 0.08517 0.56 0.574
## Year1998 -0.10746 0.08416 -1.28 0.202
## Year1999 -0.15435 0.08294 -1.86 0.063 .
## Year2000 0.11224 0.08232 1.36 0.173
## Year2001 0.03604 0.09751 0.37 0.712
## Year2002 -0.07172 0.07639 -0.94 0.348
## Year2003 -0.00387 0.07473 -0.05 0.959
## Year2004 0.01040 0.07511 0.14 0.890
## Year2005 0.03601 0.08045 0.45 0.654
```

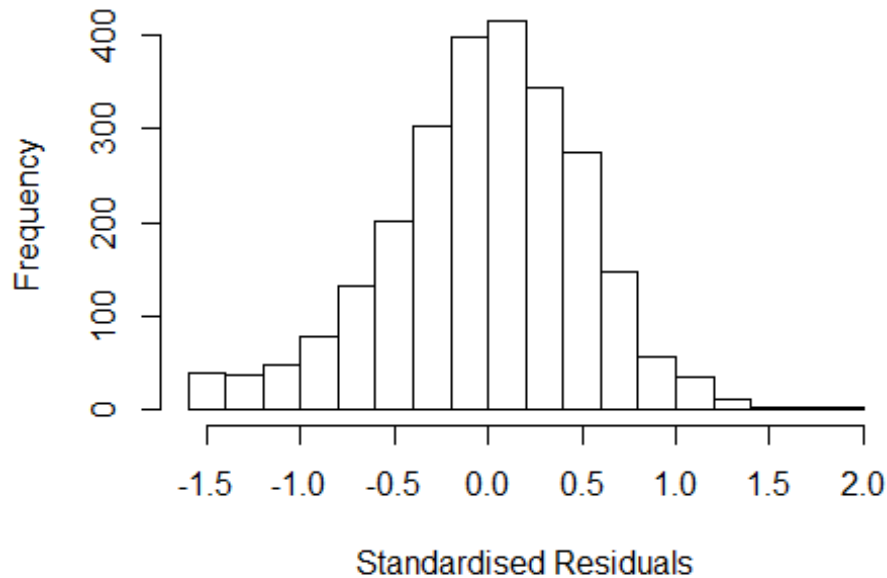


```

## Year2006          0.02201    0.07480    0.29    0.769
## Year2007          0.09964    0.07487    1.33    0.183
## Year2008          0.13157    0.07762    1.70    0.090 .
## Year2009          0.07430    0.07809    0.95    0.341
## Year2010          0.10531    0.07566    1.39    0.164
## Year2011          0.03254    0.07951    0.41    0.682
## Year2012          0.09431    0.07664    1.23    0.219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0218, Adjusted R-squared:  0.0148
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2304 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0992 0.8620 0.9500 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.055 1      1.027
## Year              1.055 16      1.002

```

## Residuals from first author



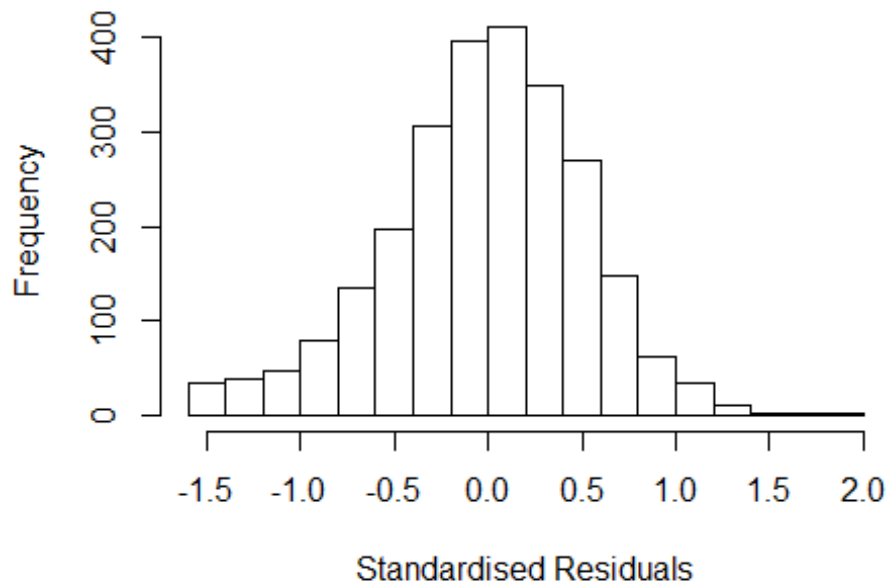
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5027 -0.3169 0.0139 0.3317 1.8955
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.37533 0.06628 20.75 <2e-16 ***
## FirstAuthorFemale1 -0.05508 0.02411 -2.28 0.022 *
## Year1997 0.04861 0.08522 0.57 0.568
## Year1998 -0.10837 0.08408 -1.29 0.198
## Year1999 -0.15530 0.08291 -1.87 0.061 .
## Year2000 0.11128 0.08226 1.35 0.176
## Year2001 0.03662 0.09740 0.38 0.707
## Year2002 -0.07293 0.07628 -0.96 0.339
## Year2003 -0.00468 0.07471 -0.06 0.950
## Year2004 0.00928 0.07508 0.12 0.902
## Year2005 0.03530 0.08042 0.44 0.661
## Year2006 0.02126 0.07475 0.28 0.776
```

```

## Year2007          0.09854    0.07481    1.32    0.188
## Year2008          0.12740    0.07715    1.65    0.099 .
## Year2009          0.07250    0.07801    0.93    0.353
## Year2010          0.10503    0.07568    1.39    0.165
## Year2011          0.03117    0.07951    0.39    0.695
## Year2012          0.09275    0.07655    1.21    0.226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0215, Adjusted R-squared:  0.0148
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 215 weights are ~= 1. The remaining 2309 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0975 0.8640 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.076 1          1.037
## Year              1.076 16          1.002

```

## Residuals from last author



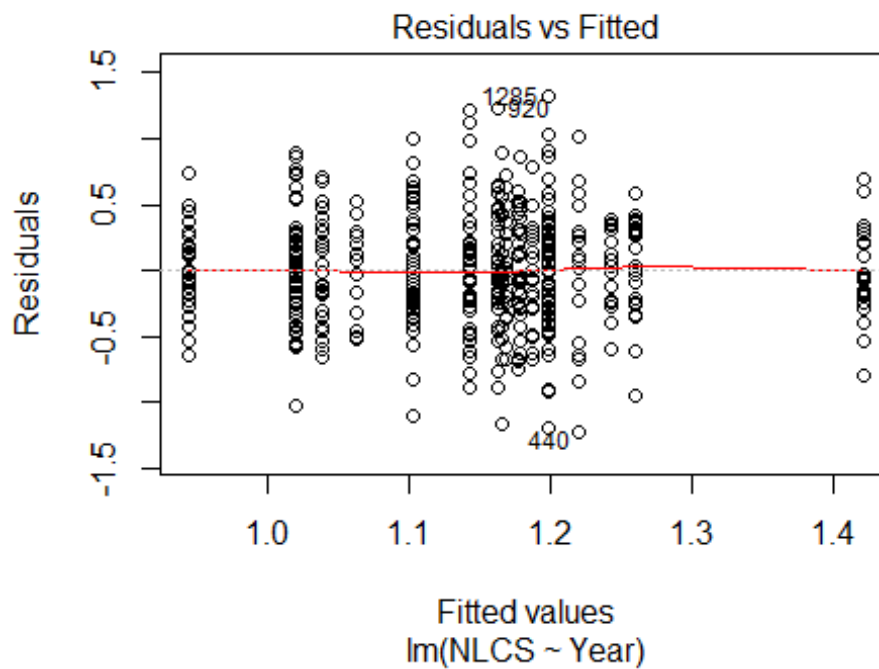
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5066 -0.3153  0.0164  0.3260  1.8829
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.37280    0.06617   20.75  <2e-16 ***
## LastAuthorFemale1 -0.05041    0.02429   -2.08   0.038 *
## Year1997         0.04911    0.08520    0.58   0.564
## Year1998        -0.10751    0.08415   -1.28   0.202
## Year1999        -0.15252    0.08299   -1.84   0.066 .
## Year2000         0.11337    0.08221    1.38   0.168
## Year2001         0.03686    0.09790    0.38   0.707
## Year2002        -0.06955    0.07640   -0.91   0.363
## Year2003        -0.00255    0.07481   -0.03   0.973
## Year2004         0.00990    0.07523    0.13   0.895
## Year2005         0.03679    0.08044    0.46   0.647
## Year2006         0.01963    0.07483    0.26   0.793
```

```

## Year2007      0.09852    0.07494    1.31    0.189
## Year2008      0.13379    0.07759    1.72    0.085 .
## Year2009      0.07432    0.07806    0.95    0.341
## Year2010      0.10473    0.07575    1.38    0.167
## Year2011      0.03278    0.07960    0.41    0.681
## Year2012      0.09580    0.07669    1.25    0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.0145
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 209 weights are ~= 1. The remaining 2315 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.103  0.863  0.952  0.895  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2524"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2302"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   58   56   57   55   56   48   37   42   36   49   64   74   68   79
## 2011 2012
##   85   96
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   17   22   24   21   16   17   18   14   19   20   26   30   43   41   52
## 2011 2012

```

```
## 53 64
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 19 21 18 14 15 15 13 18 19 25 23 41 39 49
## 2011 2012
## 47 59
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, df = 16, p-value = 0.2
```



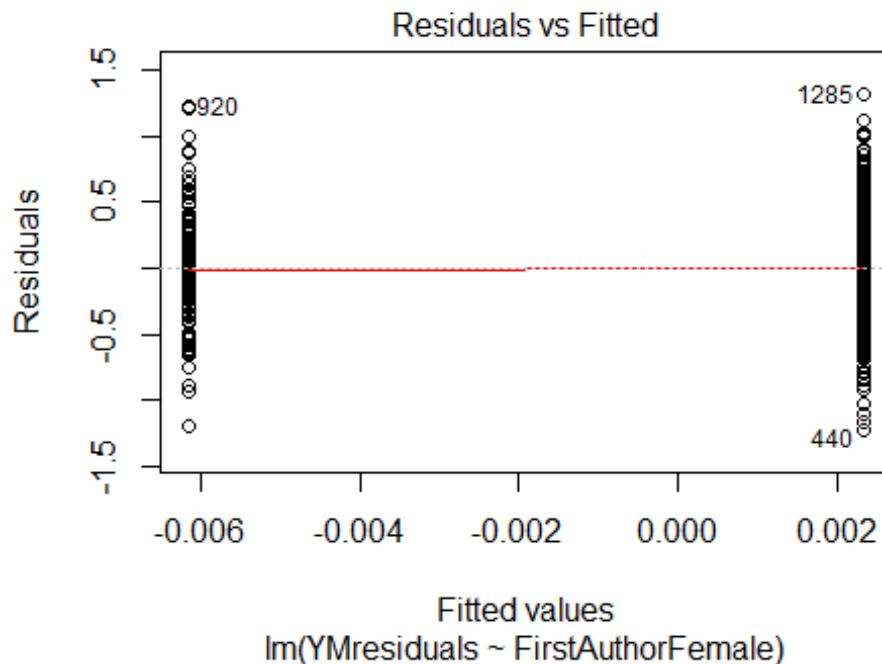
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.099, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 3.60014528450949"
## [1] "Male first author team size 2018 geometric mean: 2.92733938956642"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

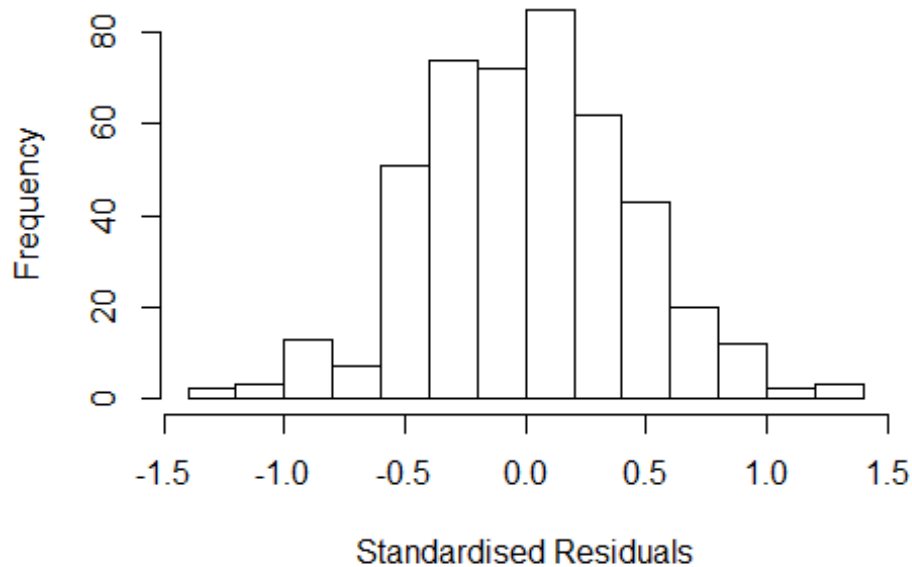
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.49400889557082"
## [1] "Male last author team size 2018 geometric mean: 2.87929337313428"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 260, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.197  1      1.094
## LastAuthorFemale  1.148  1      1.071
## UniqueAuthors    1.584  4      1.059
## Year              1.963 16      1.021
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36859 -0.28904  0.00391  0.28381  1.38440
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11447    0.14824    7.52 3.3e-13 ***
## FirstAuthorFemale1 -0.00742    0.04974   -0.15  0.8815
## LastAuthorFemale1 -0.02070    0.05575   -0.37  0.7106
## UniqueAuthors2    -0.01360    0.06165   -0.22  0.8256
## UniqueAuthors3     0.05410    0.06986    0.77  0.4392
## UniqueAuthors4     0.10129    0.07517    1.35  0.1785
## UniqueAuthors5     0.18488    0.06930    2.67  0.0079 **
## Year1997           0.09194    0.16758    0.55  0.5835
## Year1998           0.31128    0.16974    1.83  0.0674 .
## Year1999          -0.01299    0.16430   -0.08  0.9370
```

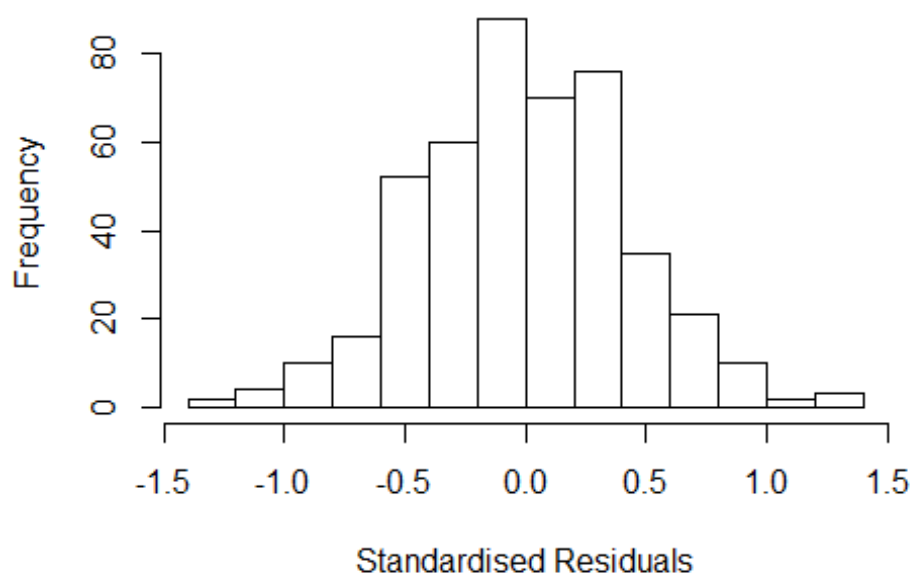


```

## Year2000      0.15062      0.16415      0.92      0.3593
## Year2001     -0.05263      0.17963     -0.29      0.7697
## Year2002      0.20003      0.24147      0.83      0.4079
## Year2003     -0.13230      0.17975     -0.74      0.4621
## Year2004      0.00974      0.17068      0.06      0.9545
## Year2005      0.01831      0.17200      0.11      0.9153
## Year2006     -0.12016      0.16304     -0.74      0.4615
## Year2007     -0.25253      0.15883     -1.59      0.1126
## Year2008     -0.04192      0.16162     -0.26      0.7955
## Year2009      0.00349      0.15614      0.02      0.9822
## Year2010     -0.15628      0.15857     -0.99      0.3249
## Year2011     -0.06646      0.16310     -0.41      0.6839
## Year2012      0.03372      0.15512      0.22      0.8280
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.434
## Multiple R-squared:  0.0862, Adjusted R-squared:  0.039
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 404 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.289  0.879  0.951  0.908  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.152 1      1.073
## LastAuthorFemale  1.122 1      1.059
## Year              1.270 16      1.007

```

## Residuals from first and last author



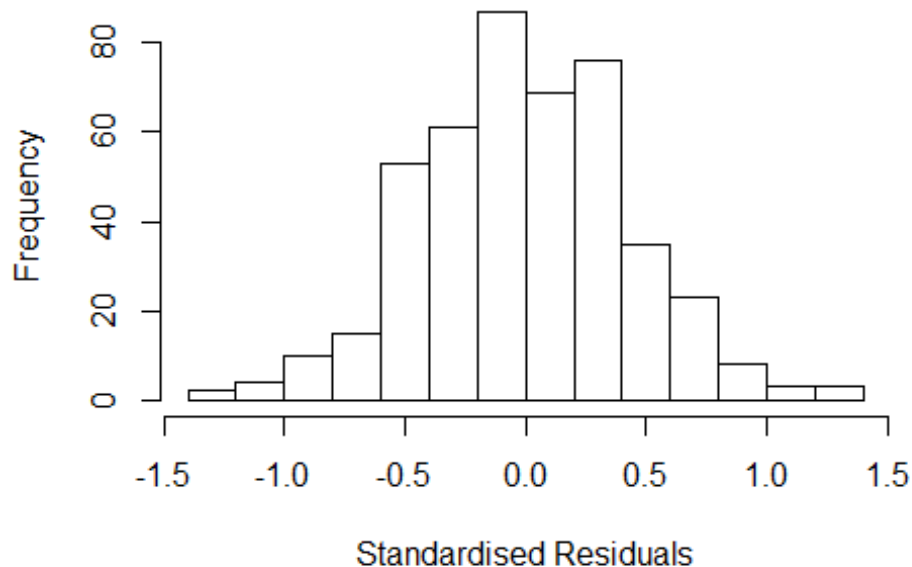
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35249 -0.28994 -0.00952 0.29790 1.31226
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12067 0.14655 7.65 1.4e-13 ***
## FirstAuthorFemale1 0.00742 0.04980 0.15 0.882
## LastAuthorFemale1 -0.01366 0.05493 -0.25 0.804
## Year1997 0.11357 0.17116 0.66 0.507
## Year1998 0.31263 0.16967 1.84 0.066 .
## Year1999 0.00143 0.16973 0.01 0.993
## Year2000 0.15663 0.16596 0.94 0.346
## Year2001 -0.00150 0.18560 -0.01 0.994
## Year2002 0.23182 0.22857 1.01 0.311
## Year2003 -0.06130 0.17839 -0.34 0.731
## Year2004 0.03859 0.17415 0.22 0.825
## Year2005 0.06841 0.17353 0.39 0.694
```

```

## Year2006      -0.10713    0.16968   -0.63    0.528
## Year2007      -0.20585    0.15993   -1.29    0.199
## Year2008       0.00648    0.16575    0.04    0.969
## Year2009       0.03285    0.15913    0.21    0.837
## Year2010      -0.10970    0.16153   -0.68    0.497
## Year2011       0.00220    0.16198    0.01    0.989
## Year2012       0.08607    0.15769    0.55    0.585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.445
## Multiple R-squared:  0.0645, Adjusted R-squared:  0.0254
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 410 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.883   0.952   0.912   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.143 1      1.069
## Year              1.143 16      1.004

```

## Residuals from first author



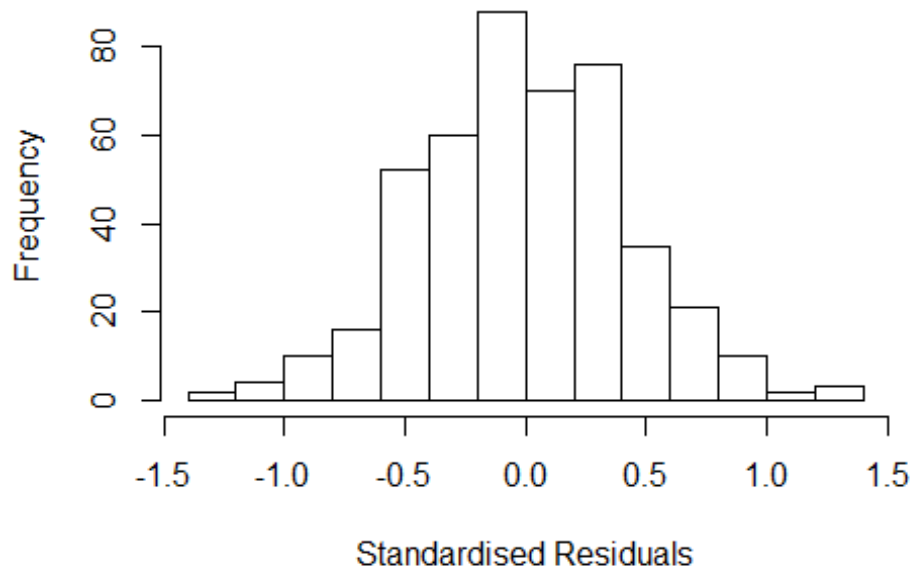
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.34958 -0.28784 -0.00811  0.29933  1.31309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.118774   0.146672   7.63 1.5e-13 ***
## FirstAuthorFemale1 0.006332   0.049613   0.13  0.898
## Year1997         0.113480   0.171527   0.66  0.509
## Year1998         0.313042   0.169603   1.85  0.066 .
## Year1999         0.001963   0.169965   0.01  0.991
## Year2000         0.157838   0.166186   0.95  0.343
## Year2001        -0.002547   0.185671  -0.01  0.989
## Year2002         0.230811   0.227842   1.01  0.312
## Year2003        -0.059109   0.178469  -0.33  0.741
## Year2004         0.038419   0.174482   0.22  0.826
## Year2005         0.068633   0.173865   0.39  0.693
## Year2006        -0.107031   0.169923  -0.63  0.529
```

```

## Year2007      -0.204971    0.160194    -1.28    0.201
## Year2008      0.006296    0.166114     0.04    0.970
## Year2009      0.033734    0.159426     0.21    0.833
## Year2010     -0.110638    0.161774    -0.68    0.494
## Year2011      0.000936    0.162173     0.01    0.995
## Year2012      0.087134    0.158006     0.55    0.582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.445
## Multiple R-squared:  0.0644, Adjusted R-squared:  0.0275
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 408 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.337  0.883  0.952  0.912  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.114 1      1.056
## Year      1.114 16      1.003

```

## Residuals from last author



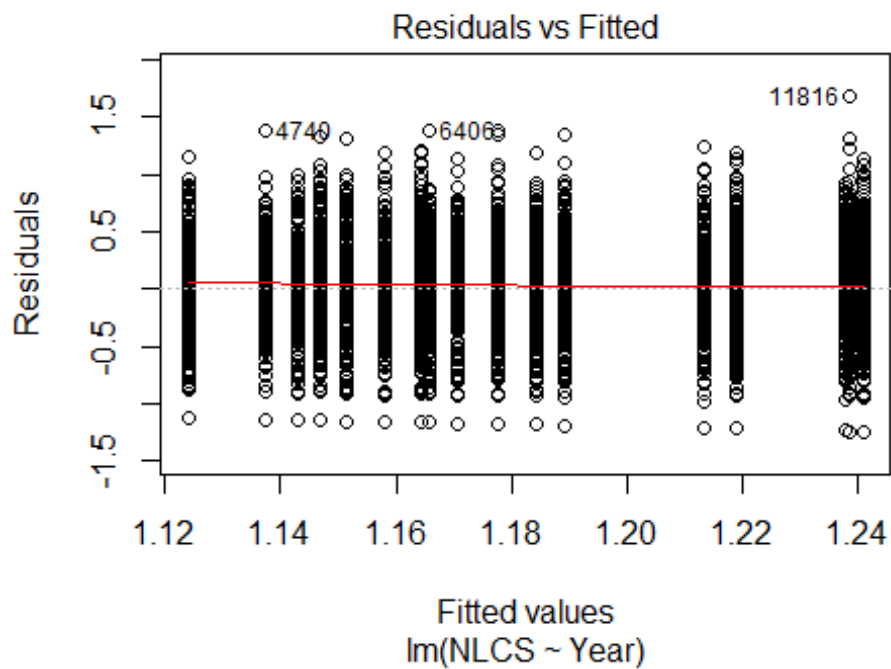
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3529 -0.2878 -0.0108 0.2982 1.3102
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.121918 0.145801 7.69 9.7e-14 ***
## LastAuthorFemale1 -0.012795 0.054766 -0.23 0.815
## Year1997 0.114414 0.171040 0.67 0.504
## Year1998 0.313176 0.169813 1.84 0.066 .
## Year1999 0.001845 0.169549 0.01 0.991
## Year2000 0.158052 0.165586 0.95 0.340
## Year2001 -0.000524 0.184996 0.00 0.998
## Year2002 0.231004 0.227420 1.02 0.310
## Year2003 -0.060559 0.178452 -0.34 0.735
## Year2004 0.039203 0.173977 0.23 0.822
## Year2005 0.068292 0.173434 0.39 0.694
## Year2006 -0.106309 0.169377 -0.63 0.531
```

```

## Year2007          -0.205509    0.159975    -1.28     0.200
## Year2008           0.006590    0.165550     0.04     0.968
## Year2009           0.033882    0.159106     0.21     0.831
## Year2010          -0.108633    0.161235    -0.67     0.501
## Year2011           0.001872    0.161720     0.01     0.991
## Year2012           0.086866    0.157538     0.55     0.582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.0645, Adjusted R-squared:  0.0276
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 411 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.333  0.883  0.952  0.912  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 449"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2303"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  485  505  489  562  535  554  527  469  417  447  490  580  556  452  483
## 2011 2012
##  510  482
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  210  214  219  266  217  203  279  258  220  257  308  367  378  298  316
## 2011 2012

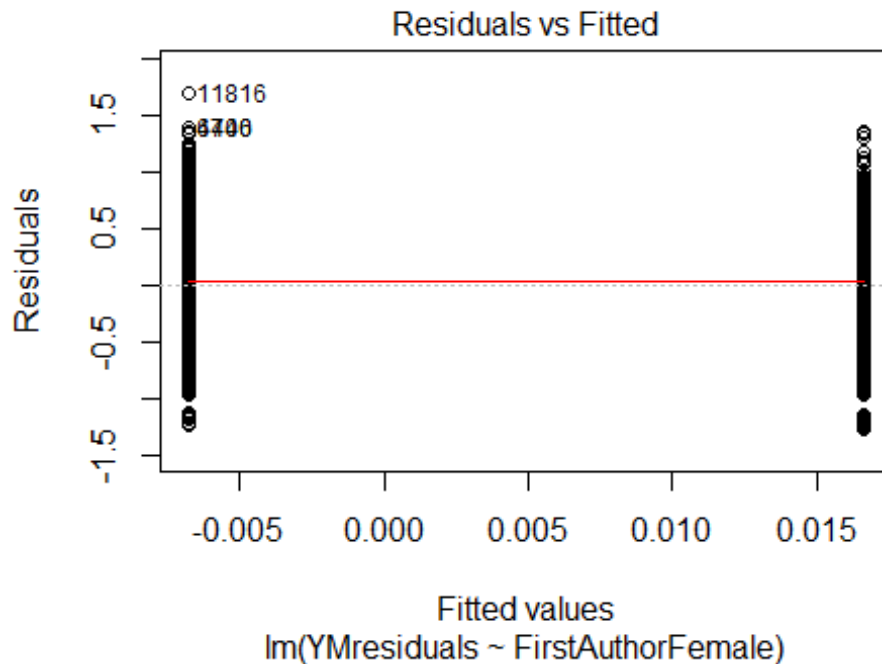
```

```
## 362 358
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 188 199 203 246 202 180 246 228 200 227 273 318 350 276 288
## 2011 2012
## 327 327
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```



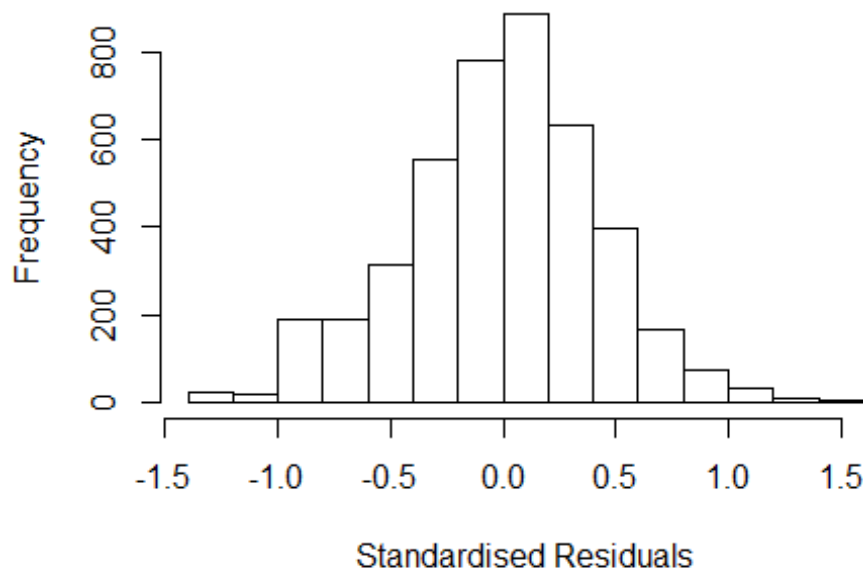
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```





```
## [1] "Female first author team size 2018 geometric mean: 3.25362336309316"
## [1] "Male first author team size 2018 geometric mean: 3.09615218323311"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.16936343801614"
## [1] "Male last author team size 2018 geometric mean: 3.1527201633153"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.050 1          1.025
## LastAuthorFemale  1.017 1          1.009
## UniqueAuthors    1.168 4          1.020
## Year             1.190 16          1.005
```

## Residuals from first and last author and team size



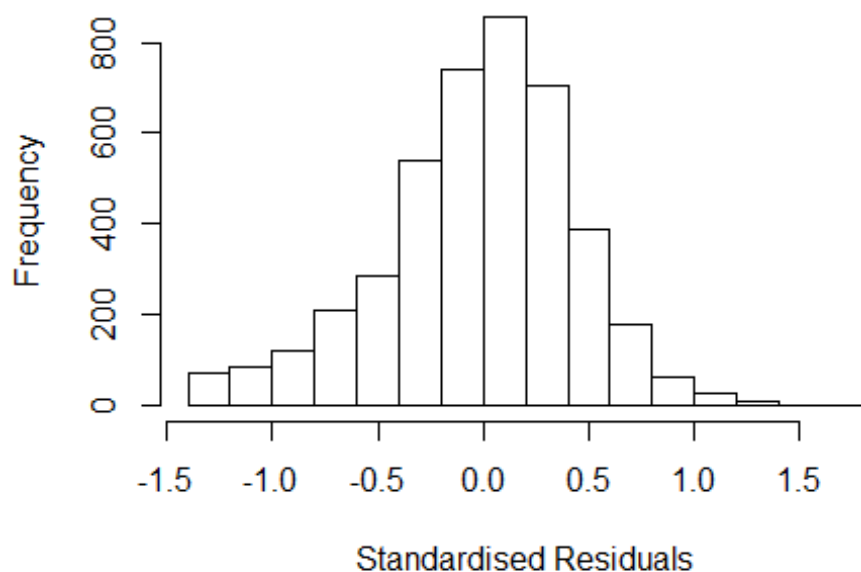
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3158 -0.2777 0.0153 0.2705 1.5418
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.993295 0.042145 23.57 <2e-16 ***
## FirstAuthorFemale1 -0.000348 0.014553 -0.02 0.98
## LastAuthorFemale1 -0.018108 0.016901 -1.07 0.28
## UniqueAuthors2 0.255466 0.025029 10.21 <2e-16 ***
## UniqueAuthors3 0.270138 0.025360 10.65 <2e-16 ***
## UniqueAuthors4 0.332977 0.026054 12.78 <2e-16 ***
## UniqueAuthors5 0.409799 0.025670 15.96 <2e-16 ***
## Year1997 -0.011803 0.051138 -0.23 0.82
## Year1998 -0.028965 0.050120 -0.58 0.56
## Year1999 -0.042670 0.048940 -0.87 0.38
```

```

## Year2000      -0.009780    0.049558   -0.20    0.84
## Year2001      -0.038718    0.051850   -0.75    0.46
## Year2002      -0.019597    0.047940   -0.41    0.68
## Year2003      -0.075354    0.048162   -1.56    0.12
## Year2004      -0.040529    0.047159   -0.86    0.39
## Year2005      -0.076099    0.046662   -1.63    0.10
## Year2006      -0.054596    0.047308   -1.15    0.25
## Year2007      -0.031351    0.045658   -0.69    0.49
## Year2008      -0.051168    0.046142   -1.11    0.27
## Year2009      -0.010075    0.046099   -0.22    0.83
## Year2010      -0.010319    0.046147   -0.22    0.82
## Year2011      -0.027240    0.046587   -0.58    0.56
## Year2012      -0.026885    0.046781   -0.57    0.57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0981, Adjusted R-squared:  0.0934
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 338 weights are ~= 1. The remaining 3940 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.861  0.948   0.894   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.040 1      1.020
## LastAuthorFemale  1.016 1      1.008
## Year              1.056 16      1.002

```

## Residuals from first and last author



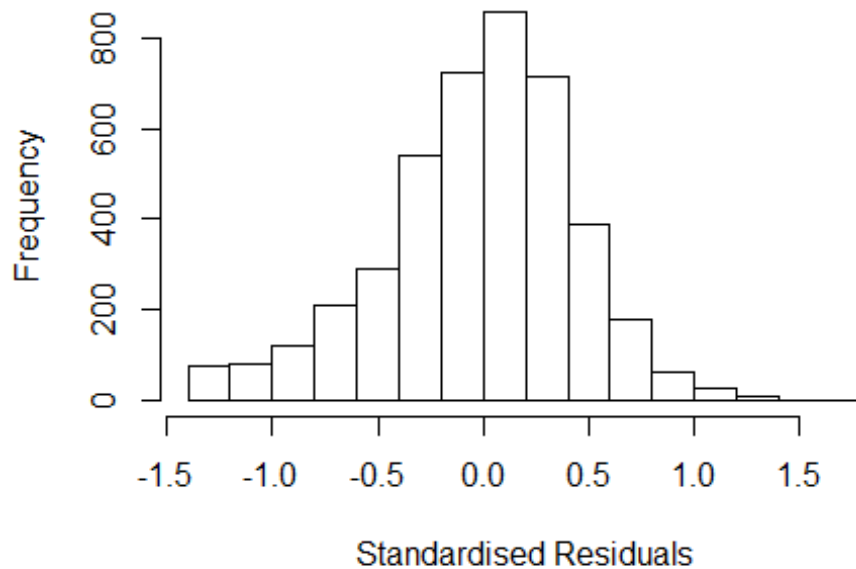
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2933 -0.2816  0.0225  0.2706  1.6666
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18309    0.04171   28.36  <2e-16 ***
## FirstAuthorFemale1  0.02752    0.01477    1.86   0.062 .
## LastAuthorFemale1 -0.03342    0.01722   -1.94   0.052 .
## Year1997         -0.01158    0.05362   -0.22   0.829
## Year1998         -0.01462    0.05164   -0.28   0.777
## Year1999         -0.03297    0.05150   -0.64   0.522
## Year2000          0.00352    0.05359    0.07   0.948
## Year2001          0.00203    0.05561    0.04   0.971
## Year2002          0.03681    0.05004    0.74   0.462
## Year2003         -0.00713    0.05090   -0.14   0.889
## Year2004          0.02463    0.04984    0.49   0.621
## Year2005         -0.00563    0.04895   -0.11   0.908
```

```

## Year2006          0.02697    0.04882    0.55    0.581
## Year2007          0.03095    0.04793    0.65    0.519
## Year2008          0.01341    0.04842    0.28    0.782
## Year2009          0.08266    0.04828    1.71    0.087 .
## Year2010          0.07168    0.04798    1.49    0.135
## Year2011          0.04995    0.04864    1.03    0.304
## Year2012          0.06835    0.04870    1.40    0.161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.00732,    Adjusted R-squared:  0.00312
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 333 weights are ~= 1. The remaining 3945 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.059  0.858  0.951  0.888  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```

## Residuals from first author



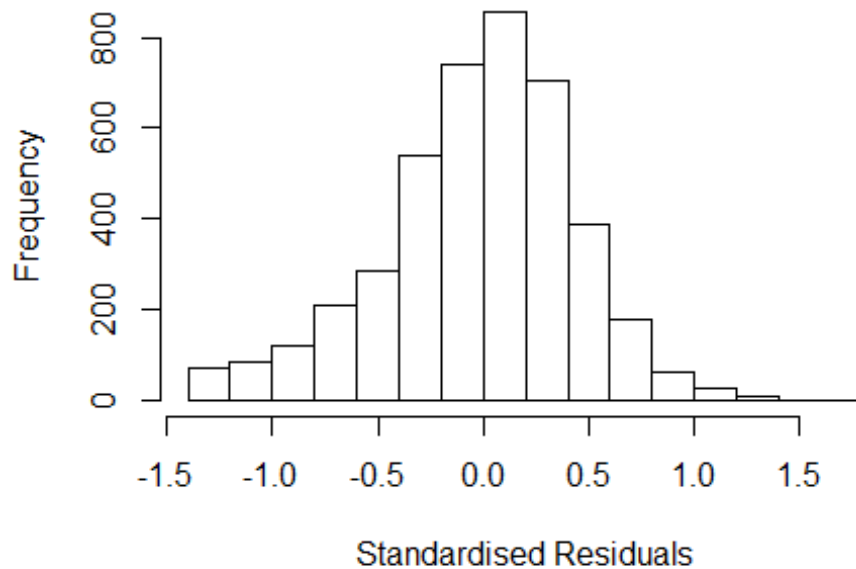
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2846 -0.2824 0.0254 0.2704 1.6713
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17620 0.04138 28.42 <2e-16 ***
## FirstAuthorFemale1 0.02341 0.01496 1.56 0.118
## Year1997 -0.00955 0.05361 -0.18 0.859
## Year1998 -0.01203 0.05159 -0.23 0.816
## Year1999 -0.02987 0.05145 -0.58 0.562
## Year2000 0.00422 0.05366 0.08 0.937
## Year2001 0.00443 0.05555 0.08 0.936
## Year2002 0.04001 0.05002 0.80 0.424
## Year2003 -0.00617 0.05087 -0.12 0.904
## Year2004 0.02750 0.04979 0.55 0.581
## Year2005 -0.00267 0.04891 -0.05 0.956
## Year2006 0.02799 0.04885 0.57 0.567
```

```

## Year2007          0.03200      0.04789      0.67      0.504
## Year2008          0.01533      0.04840      0.32      0.751
## Year2009          0.08504      0.04820      1.76      0.078 .
## Year2010          0.07271      0.04803      1.51      0.130
## Year2011          0.05004      0.04860      1.03      0.303
## Year2012          0.07046      0.04862      1.45      0.147
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.00652,    Adjusted R-squared:  0.00256
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 337 weights are ~= 1. The remaining 3941 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0572 0.8570 0.9510 0.8880 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1      1.009
## Year              1.017 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2751 -0.2787 0.0228 0.2724 1.6574
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18759 0.04169 28.49 <2e-16 ***
## LastAuthorFemale1 -0.02877 0.01734 -1.66 0.097 .
## Year1997 -0.01140 0.05371 -0.21 0.832
## Year1998 -0.01360 0.05161 -0.26 0.792
## Year1999 -0.03240 0.05154 -0.63 0.530
## Year2000 0.00523 0.05354 0.10 0.922
## Year2001 0.00479 0.05566 0.09 0.931
## Year2002 0.03838 0.05012 0.77 0.444
## Year2003 -0.00446 0.05083 -0.09 0.930
## Year2004 0.02793 0.04980 0.56 0.575
## Year2005 -0.00225 0.04891 -0.05 0.963
## Year2006 0.03006 0.04874 0.62 0.537
```

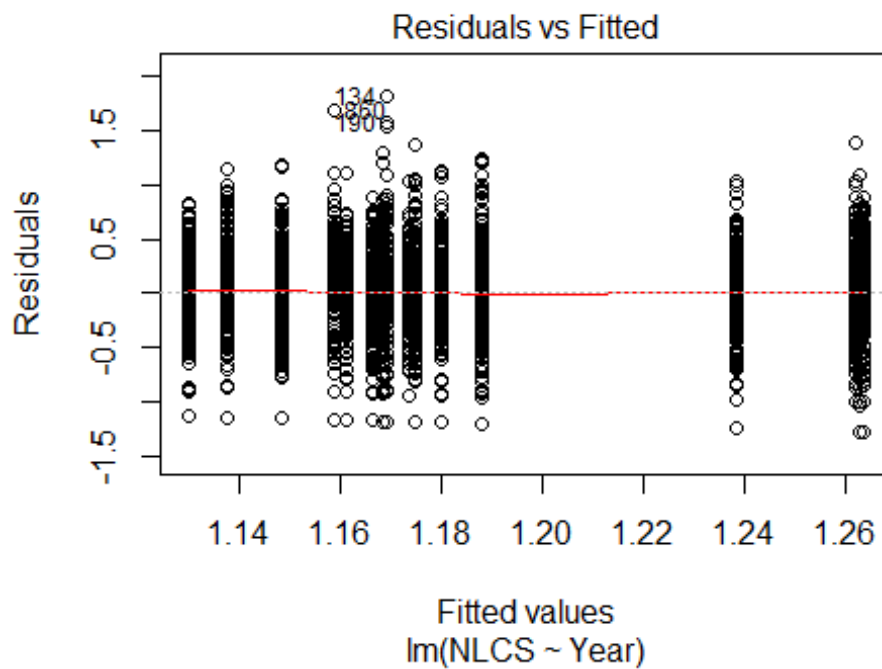


```

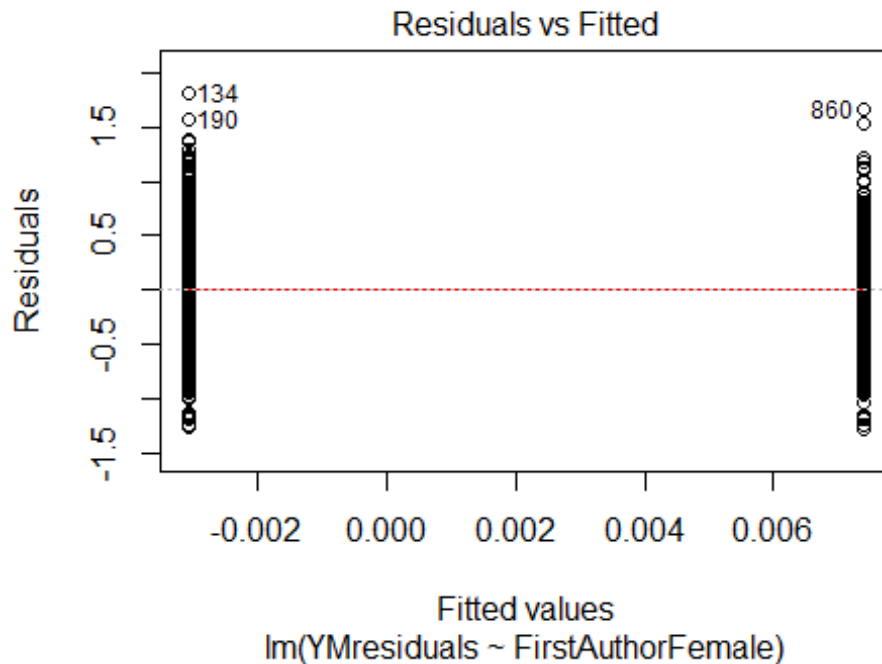
## Year2007      0.03370      0.04792      0.70      0.482
## Year2008      0.01788      0.04832      0.37      0.711
## Year2009      0.08755      0.04818      1.82      0.069 .
## Year2010      0.07473      0.04801      1.56      0.120
## Year2011      0.05335      0.04859      1.10      0.272
## Year2012      0.07298      0.04860      1.50      0.133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.00663,    Adjusted R-squared:  0.00267
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 329 weights are ~= 1. The remaining 3949 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0634 0.8580 0.9510 0.8880 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4278"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2304"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 312 352 400 397 380 409 374 320 337 284 376 368 322 310 353
## 2011 2012
## 371 413
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 136 133 176 129 122 191 168 179 172 212 223 195 193 197
## 2011 2012

```

```
## 219 270
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 99 126 117 165 123 109 174 152 156 154 190 196 173 171 174
## 2011 2012
## 189 245
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 75, df = 16, p-value = 1e-09
```

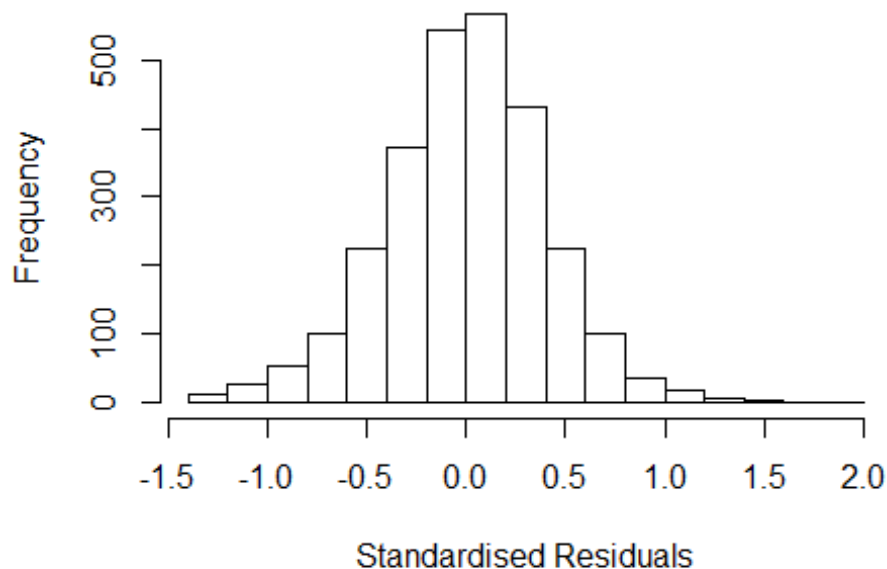


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 4.22447592907821"
## [1] "Male first author team size 2018 geometric mean: 3.87712649947285"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5800, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.71119996153673"
## [1] "Male last author team size 2018 geometric mean: 4.11491361743449"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3700, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1      1.015
## LastAuthorFemale  1.048 1      1.023
## UniqueAuthors    1.231 4      1.026
## Year             1.260 16      1.007
```

## Residuals from first and last author and team size



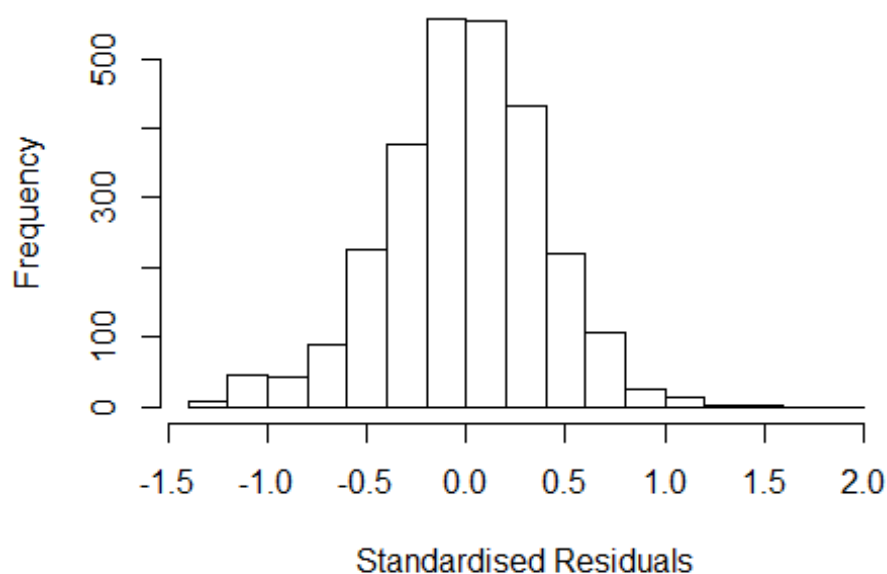
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35784 -0.24725 0.00953 0.24437 1.81067
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99296 0.06569 15.12 < 2e-16 ***
## FirstAuthorFemale1 -0.00671 0.01594 -0.42 0.674
## LastAuthorFemale1 -0.02055 0.01877 -1.09 0.274
## UniqueAuthors2 0.17037 0.03624 4.70 2.7e-06 ***
## UniqueAuthors3 0.17286 0.03588 4.82 1.5e-06 ***
## UniqueAuthors4 0.19321 0.03647 5.30 1.3e-07 ***
## UniqueAuthors5 0.24271 0.03560 6.82 1.1e-11 ***
## Year1997 0.14272 0.06858 2.08 0.038 *
## Year1998 0.01934 0.06976 0.28 0.782
## Year1999 0.04432 0.06628 0.67 0.504
```

```

## Year2000          0.12151      0.06609      1.84      0.066 .
## Year2001          0.04027      0.06723      0.60      0.549
## Year2002          0.03891      0.06408      0.61      0.544
## Year2003          0.01406      0.06260      0.22      0.822
## Year2004          0.00662      0.06188      0.11      0.915
## Year2005         -0.01767      0.06337     -0.28      0.780
## Year2006          0.01119      0.06185      0.18      0.856
## Year2007         -0.00108      0.06110     -0.02      0.986
## Year2008          0.06157      0.06257      0.98      0.325
## Year2009         -0.02675      0.06341     -0.42      0.673
## Year2010          0.07639      0.06258      1.22      0.222
## Year2011          0.00685      0.06413      0.11      0.915
## Year2012         -0.02019      0.06167     -0.33      0.743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.0384, Adjusted R-squared:  0.0305
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 44 is an outlier with |weight| = 0 ( < 3.7e-05);
## 212 weights are ~= 1. The remaining 2500 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0306 0.8670 0.9530 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.69e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1          1.013
## LastAuthorFemale  1.028 1          1.014
## Year              1.050 16          1.002

```

## Residuals from first and last author



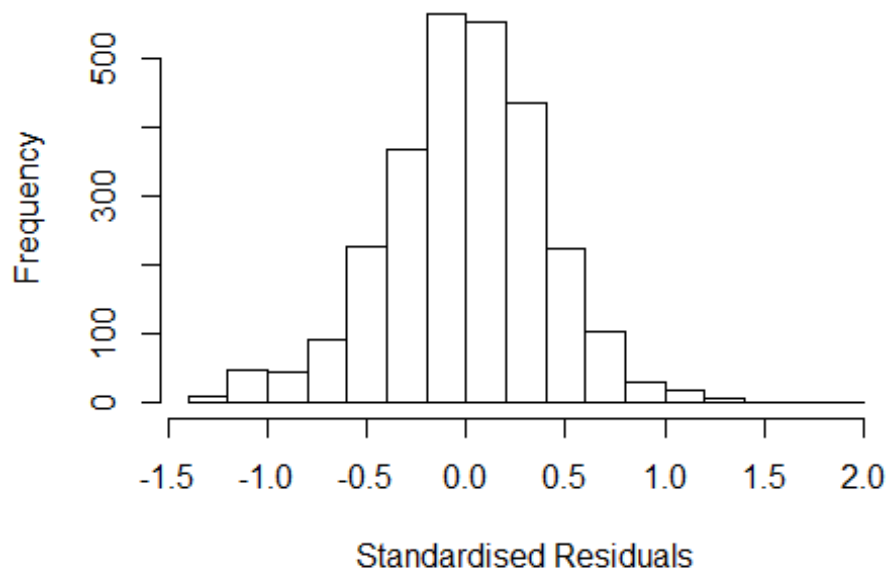
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29064 -0.24769  0.00347  0.25147  1.82148
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.152516   0.053782   21.43  <2e-16 ***
## FirstAuthorFemale1  0.005526   0.016142    0.34   0.732
## LastAuthorFemale1 -0.020243   0.018969   -1.07   0.286
## Year1997         0.138123   0.066915    2.06   0.039 *
## Year1998         0.009688   0.068357    0.14   0.887
## Year1999         0.041644   0.064424    0.65   0.518
## Year2000         0.124249   0.064980    1.91   0.056 .
## Year2001         0.052341   0.066693    0.78   0.433
## Year2002         0.040009   0.062047    0.64   0.519
## Year2003         0.032513   0.060295    0.54   0.590
## Year2004         0.023177   0.059904    0.39   0.699
## Year2005         0.002483   0.061696    0.04   0.968
```

```

## Year2006          0.025919    0.060106    0.43    0.666
## Year2007          0.011217    0.059358    0.19    0.850
## Year2008          0.081173    0.060816    1.33    0.182
## Year2009          0.000226    0.061510    0.00    0.997
## Year2010          0.105945    0.060606    1.75    0.081 .
## Year2011          0.037361    0.062567    0.60    0.550
## Year2012          0.013814    0.059427    0.23    0.816
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00419
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 44 is an outlier with |weight| = 0 ( < 3.7e-05);
## 222 weights are ~ = 1. The remaining 2490 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0125 0.8700 0.9530 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1          1.012
## Year              1.024 16          1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28829 -0.24794 0.00231 0.25273 1.82457
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.149428 0.053749 21.39 <2e-16 ***
## FirstAuthorFemale1 0.004507 0.016130 0.28 0.780
## Year1997 0.138860 0.066961 2.07 0.038 *
## Year1998 0.008864 0.068496 0.13 0.897
## Year1999 0.041611 0.064549 0.64 0.519
## Year2000 0.122831 0.065080 1.89 0.059 .
## Year2001 0.052821 0.066733 0.79 0.429
## Year2002 0.039019 0.062178 0.63 0.530
## Year2003 0.032514 0.060388 0.54 0.590
## Year2004 0.022755 0.060020 0.38 0.705
## Year2005 0.002618 0.061824 0.04 0.966
## Year2006 0.025448 0.060236 0.42 0.673
```

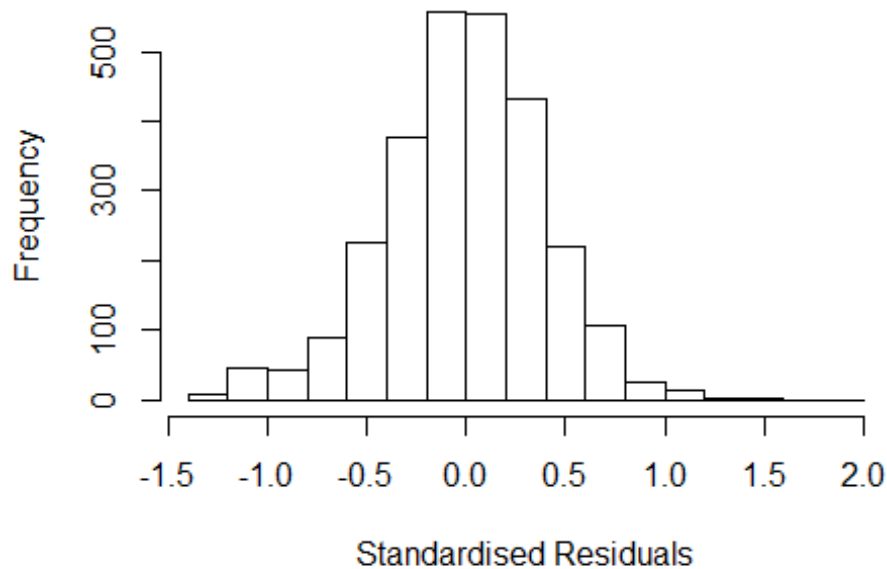


```

## Year2007      0.010562    0.059494    0.18    0.859
## Year2008      0.080794    0.060962    1.33    0.185
## Year2009      0.000501    0.061645    0.01    0.994
## Year2010      0.105343    0.060739    1.73    0.083 .
## Year2011      0.037834    0.062606    0.60    0.546
## Year2012      0.013839    0.059533    0.23    0.816
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.0042
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 44 is an outlier with |weight| = 0 ( < 3.7e-05);
## 215 weights are ~= 1. The remaining 2497 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0113 0.8690 0.9530 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.026 1          1.013
## Year            1.026 16          1.001

```

## Residuals from last author



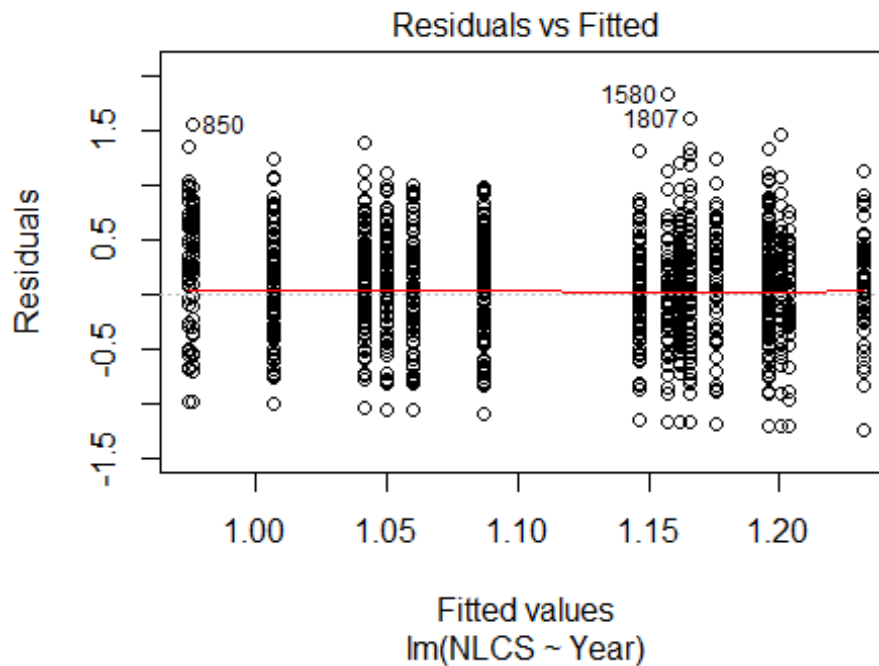
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2916 -0.2466 0.0035 0.2502 1.8206
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15344 0.05379 21.44 <2e-16 ***
## LastAuthorFemale1 -0.01985 0.01895 -1.05 0.295
## Year1997 0.13821 0.06698 2.06 0.039 *
## Year1998 0.01008 0.06832 0.15 0.883
## Year1999 0.04163 0.06447 0.65 0.519
## Year2000 0.12467 0.06497 1.92 0.055 .
## Year2001 0.05297 0.06666 0.79 0.427
## Year2002 0.04051 0.06205 0.65 0.514
## Year2003 0.03324 0.06027 0.55 0.581
## Year2004 0.02375 0.05988 0.40 0.692
## Year2005 0.00314 0.06167 0.05 0.959
## Year2006 0.02694 0.06005 0.45 0.654
```

```

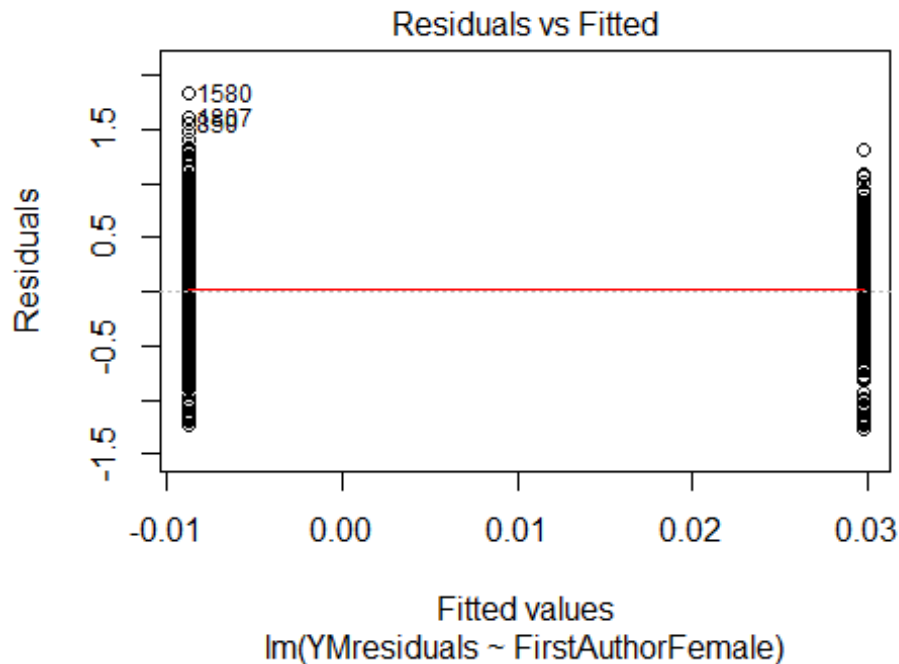
## Year2007      0.01253      0.05920      0.21      0.832
## Year2008      0.08206      0.06076      1.35      0.177
## Year2009      0.00133      0.06142      0.02      0.983
## Year2010      0.10697      0.06054      1.77      0.077 .
## Year2011      0.03822      0.06247      0.61      0.541
## Year2012      0.01472      0.05938      0.25      0.804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.378
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00452
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 44 is an outlier with |weight| = 0 ( < 3.7e-05);
## 227 weights are ~= 1. The remaining 2485 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0114 0.8700 0.9530 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2713"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2305"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 202 239 259 242 260 253 200 190 171 195 232 245 225 251 264
## 2011 2012
## 267 303
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 60 62 60 64 59 66 58 48 101 110 111 105 115 117

```

```
## 2011 2012
## 131 148
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 53 54 58 58 58 49 60 50 42 92 97 88 91 102 102
## 2011 2012
## 114 132
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 46, df = 16, p-value = 1e-04
```

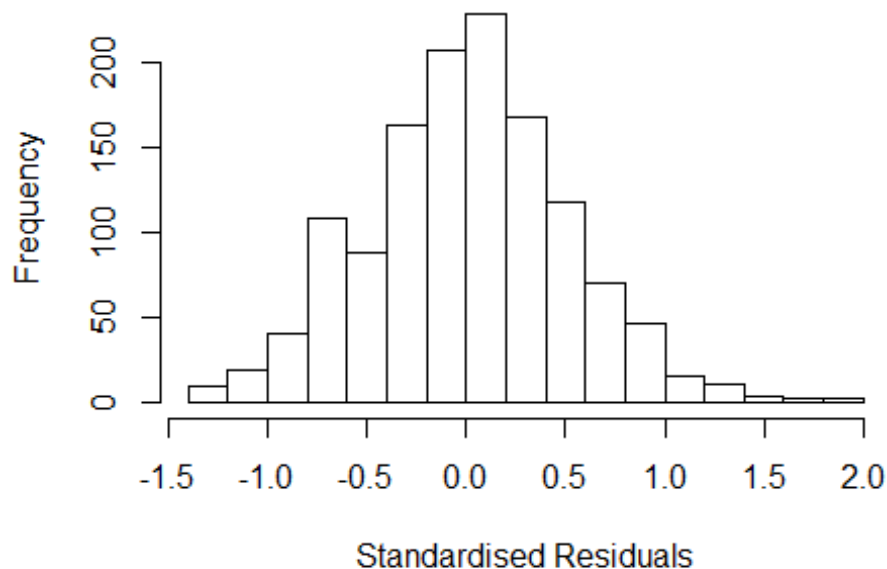


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 8.3, df = 1, p-value = 0.004
```



```
## [1] "Female first author team size 2018 geometric mean: 3.60752655054272"
## [1] "Male first author team size 2018 geometric mean: 3.32680129614556"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.2646035731464"
## [1] "Male last author team size 2018 geometric mean: 3.45856771624753"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.067 1          1.033
## LastAuthorFemale  1.101 1          1.049
## UniqueAuthors    1.313 4          1.035
## Year             1.414 16          1.011
```

## Residuals from first and last author and team size



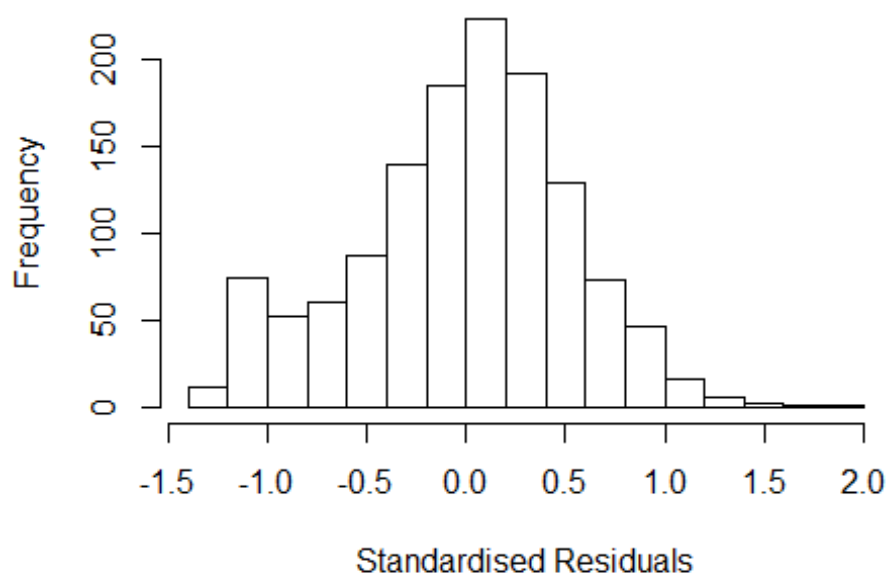
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3491 -0.3285  0.0123  0.3251  1.9407
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.66942    0.09783   6.84 1.2e-11 ***
## FirstAuthorFemale1 -0.03371    0.03214  -1.05  0.295
## LastAuthorFemale1  0.02388    0.03553   0.67  0.502
## UniqueAuthors2    0.47932    0.04797   9.99 < 2e-16 ***
## UniqueAuthors3    0.51390    0.04855  10.59 < 2e-16 ***
## UniqueAuthors4    0.52465    0.05208  10.07 < 2e-16 ***
## UniqueAuthors5    0.57688    0.04913  11.74 < 2e-16 ***
## Year1997          0.16577    0.11918   1.39  0.164
## Year1998          0.16577    0.11186   1.48  0.139
## Year1999         -0.00442    0.12155  -0.04  0.971
```

```

## Year2000      0.22306      0.11400      1.96      0.051 .
## Year2001      0.08555      0.12242      0.70      0.485
## Year2002      0.14985      0.12586      1.19      0.234
## Year2003      0.03454      0.10955      0.32      0.753
## Year2004      0.08321      0.13143      0.63      0.527
## Year2005      0.01829      0.10742      0.17      0.865
## Year2006     -0.00485      0.10677     -0.05      0.964
## Year2007      0.03084      0.10483      0.29      0.769
## Year2008     -0.03553      0.10570     -0.34      0.737
## Year2009      0.07590      0.10165      0.75      0.455
## Year2010     -0.00607      0.10464     -0.06      0.954
## Year2011     -0.04867      0.10275     -0.47      0.636
## Year2012      0.08365      0.10256      0.82      0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.481
## Multiple R-squared:  0.18, Adjusted R-squared:  0.166
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 109 weights are ~= 1. The remaining 1191 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0666 0.8510 0.9500 0.9000 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.064 1      1.031
## LastAuthorFemale 1.120 1      1.058
## Year      1.190 16      1.005

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3451 -0.3407 0.0388 0.3305 1.8134
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9200 0.1276 7.21 9.4e-13 ***
## FirstAuthorFemale1 0.0273 0.0335 0.81 0.416
## LastAuthorFemale1 0.0533 0.0373 1.43 0.153
## Year1997 0.2800 0.1548 1.81 0.071 .
## Year1998 0.2655 0.1440 1.84 0.066 .
## Year1999 0.0394 0.1688 0.23 0.815
## Year2000 0.3444 0.1447 2.38 0.017 *
## Year2001 0.1932 0.1462 1.32 0.186
## Year2002 0.2040 0.1576 1.29 0.196
## Year2003 0.2441 0.1389 1.76 0.079 .
## Year2004 0.2155 0.1629 1.32 0.186
## Year2005 0.1536 0.1389 1.11 0.269
```

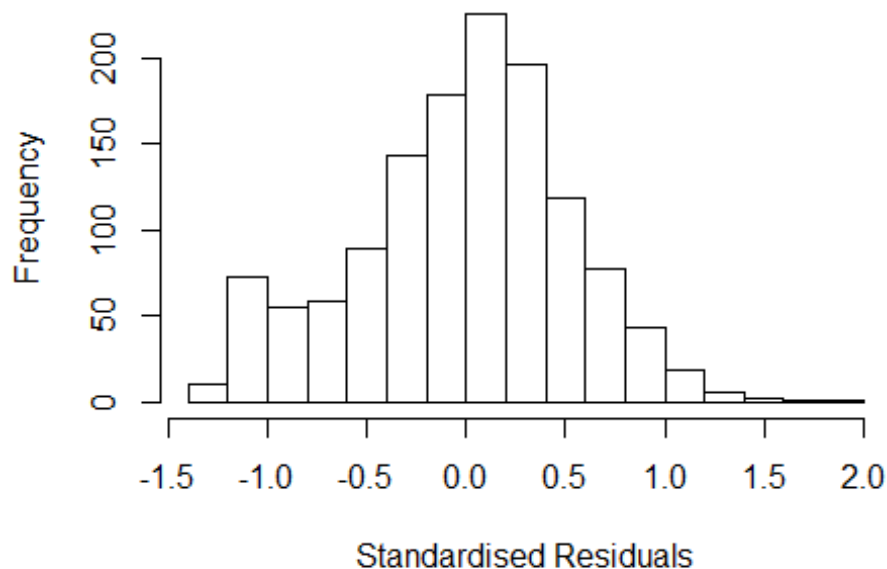


```

## Year2006          0.1304      0.1360      0.96      0.338
## Year2007          0.1465      0.1354      1.08      0.280
## Year2008          0.1222      0.1385      0.88      0.378
## Year2009          0.2352      0.1323      1.78      0.076 .
## Year2010          0.2019      0.1344      1.50      0.133
## Year2011          0.1072      0.1348      0.80      0.427
## Year2012          0.2484      0.1327      1.87      0.062 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.00895
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1186 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.152  0.839   0.950   0.891   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.068 1      1.034
## Year              1.068 16      1.002

```

## Residuals from first author



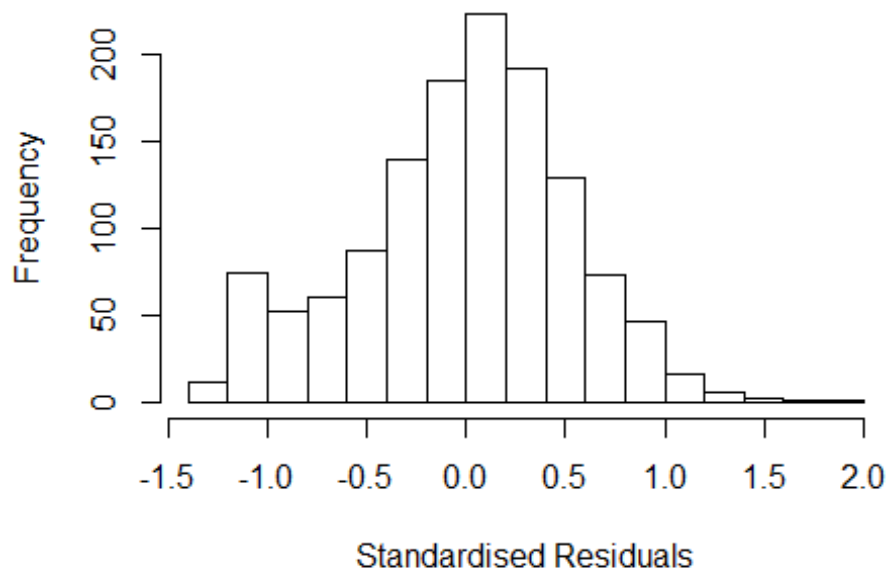
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3096 -0.3422 0.0429 0.3277 1.8595
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9308 0.1259 7.39 2.6e-13 ***
## FirstAuthorFemale1 0.0361 0.0339 1.07 0.286
## Year1997 0.2713 0.1537 1.76 0.078 .
## Year1998 0.2592 0.1432 1.81 0.071 .
## Year1999 0.0362 0.1679 0.22 0.829
## Year2000 0.3426 0.1440 2.38 0.018 *
## Year2001 0.1897 0.1451 1.31 0.191
## Year2002 0.2013 0.1572 1.28 0.201
## Year2003 0.2323 0.1369 1.70 0.090 .
## Year2004 0.2085 0.1621 1.29 0.199
## Year2005 0.1509 0.1381 1.09 0.275
## Year2006 0.1277 0.1351 0.95 0.345
```

```

## Year2007          0.1423      0.1345      1.06      0.290
## Year2008          0.1215      0.1380      0.88      0.379
## Year2009          0.2322      0.1313      1.77      0.077 .
## Year2010          0.2005      0.1337      1.50      0.134
## Year2011          0.1031      0.1339      0.77      0.442
## Year2012          0.2456      0.1319      1.86      0.063 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0211, Adjusted R-squared:  0.00808
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1196 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.129  0.846  0.950  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.127 1      1.061
## Year              1.127 16      1.004

```

## Residuals from last author



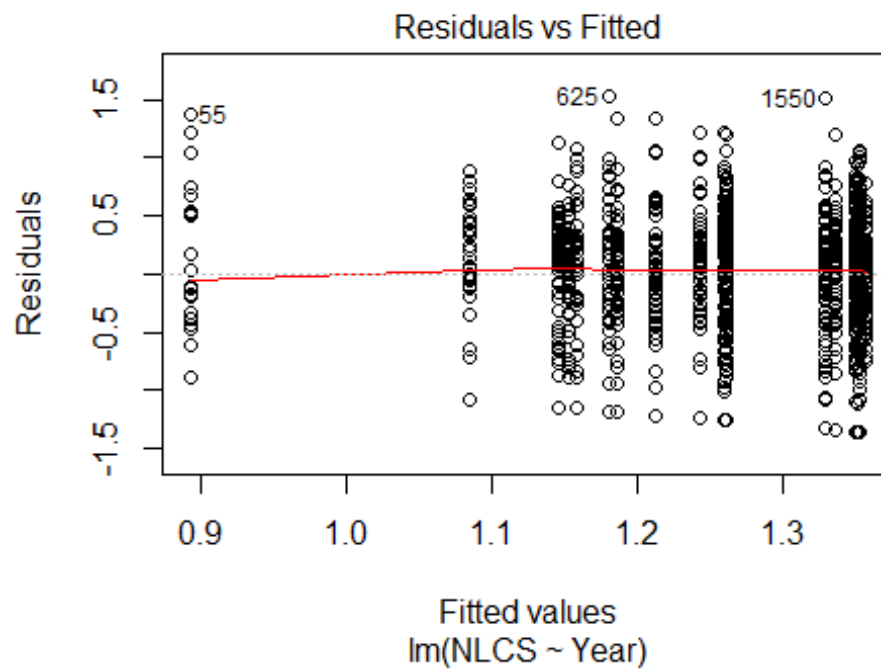
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3265 -0.3379 0.0376 0.3351 1.8017
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9278 0.1259 7.37 3.1e-13 ***
## LastAuthorFemale1 0.0590 0.0377 1.56 0.118
## Year1997 0.2780 0.1547 1.80 0.073 .
## Year1998 0.2614 0.1432 1.82 0.068 .
## Year1999 0.0337 0.1681 0.20 0.841
## Year2000 0.3397 0.1437 2.36 0.018 *
## Year2001 0.1914 0.1459 1.31 0.190
## Year2002 0.1997 0.1571 1.27 0.204
## Year2003 0.2431 0.1385 1.76 0.079 .
## Year2004 0.2160 0.1629 1.33 0.185
## Year2005 0.1500 0.1384 1.08 0.279
## Year2006 0.1292 0.1358 0.95 0.342
```

```

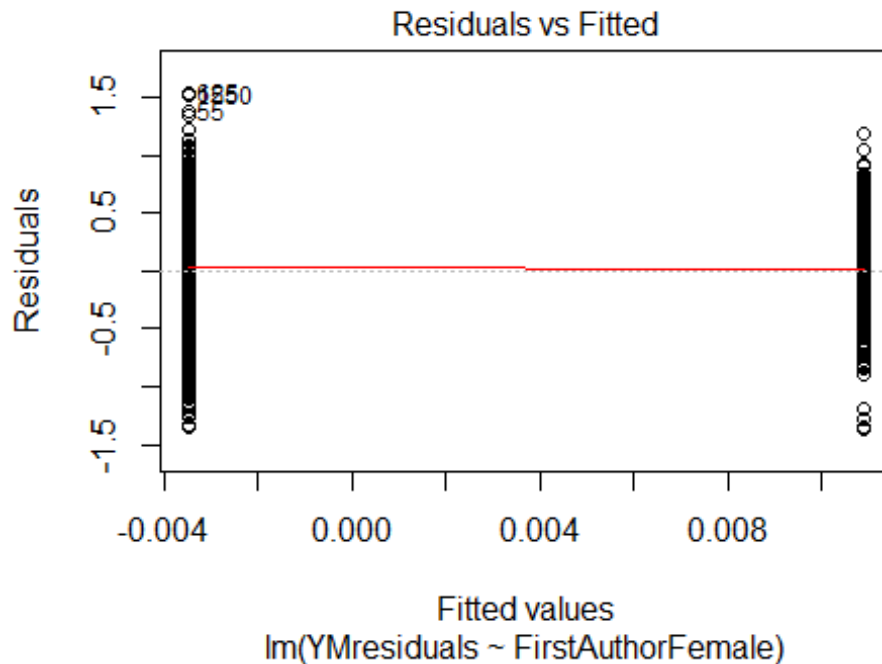
## Year2007          0.1447      0.1351      1.07      0.285
## Year2008          0.1205      0.1382      0.87      0.383
## Year2009          0.2322      0.1318      1.76      0.078 .
## Year2010          0.2003      0.1340      1.49      0.135
## Year2011          0.1056      0.1345      0.78      0.433
## Year2012          0.2467      0.1324      1.86      0.063 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.495
## Multiple R-squared:  0.0221, Adjusted R-squared:  0.00911
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1182 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.158  0.838  0.948  0.890  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1300"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2306"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   70   68   97   83   99   69   78   82   84  101  111  121  112  154
## 2011 2012
##  169  156
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   26   39   37   58   43   39   48   53   55   59   70   75   78   77  107
## 2011 2012

```

```
## 130 112
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 25 36 35 56 39 36 46 51 53 52 60 69 71 67 97
## 2011 2012
## 117 106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 44, df = 16, p-value = 2e-04
```

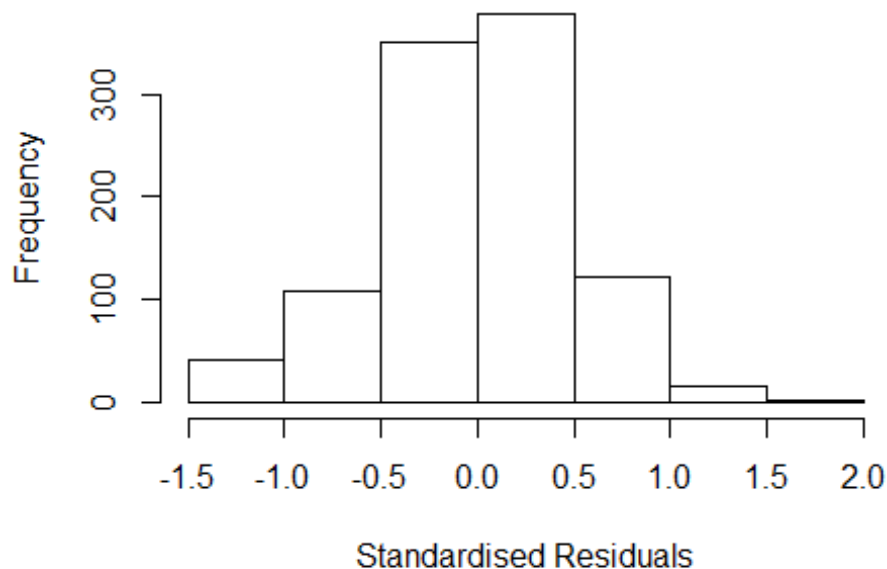


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.23329297302905"
## [1] "Male first author team size 2018 geometric mean: 2.39155585101176"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.77891435868048"
## [1] "Male last author team size 2018 geometric mean: 2.56131149832655"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.121 1      1.059
## LastAuthorFemale  1.175 1      1.084
## UniqueAuthors    1.421 4      1.045
## Year              1.481 16     1.012
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43290 -0.29348 0.00861 0.31336 1.67201
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.73107 0.14115 5.18 2.7e-07 ***
## FirstAuthorFemale1 -0.00394 0.03592 -0.11 0.91273
## LastAuthorFemale1 -0.04662 0.03976 -1.17 0.24125
## UniqueAuthors2 0.19867 0.04169 4.77 2.2e-06 ***
## UniqueAuthors3 0.20503 0.04911 4.18 3.2e-05 ***
## UniqueAuthors4 0.17707 0.05575 3.18 0.00154 **
## UniqueAuthors5 0.31545 0.04578 6.89 9.8e-12 ***
## Year1997 0.50885 0.18177 2.80 0.00522 **
## Year1998 0.26714 0.16723 1.60 0.11048
## Year1999 0.30120 0.15688 1.92 0.05515 .
```

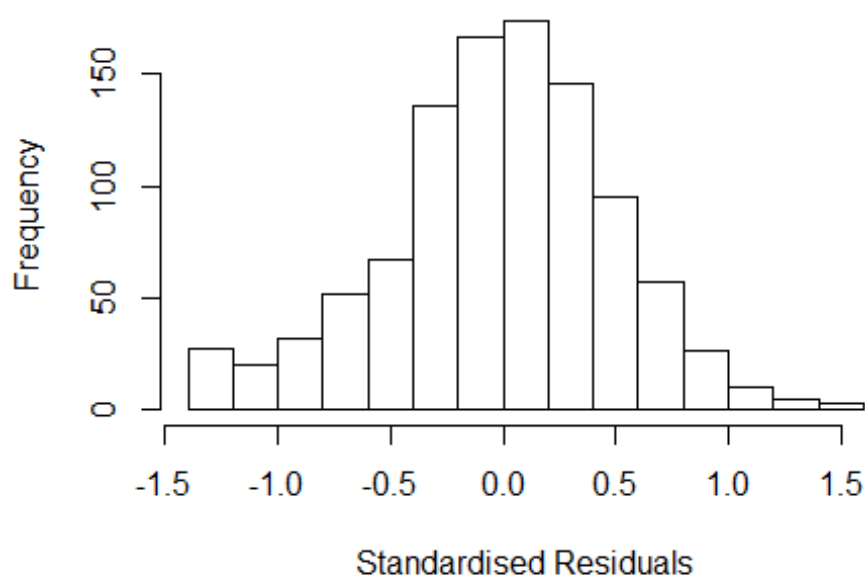


```

## Year2000      0.31168    0.15342    2.03  0.04247 *
## Year2001      0.36255    0.17561    2.06  0.03922 *
## Year2002      0.36202    0.16757    2.16  0.03098 *
## Year2003      0.31453    0.15218    2.07  0.03901 *
## Year2004      0.43251    0.16104    2.69  0.00736 **
## Year2005      0.34693    0.15900    2.18  0.02935 *
## Year2006      0.47900    0.15596    3.07  0.00219 **
## Year2007      0.47960    0.14996    3.20  0.00143 **
## Year2008      0.44592    0.15552    2.87  0.00423 **
## Year2009      0.49217    0.14826    3.32  0.00093 ***
## Year2010      0.52870    0.14830    3.57  0.00038 ***
## Year2011      0.50072    0.14862    3.37  0.00078 ***
## Year2012      0.37330    0.14904    2.50  0.01241 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0813
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 91 weights are ~= 1. The remaining 925 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.136  0.858  0.948  0.891  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.105 1      1.051
## LastAuthorFemale  1.123 1      1.060
## Year              1.162 16      1.005

```

## Residuals from first and last author



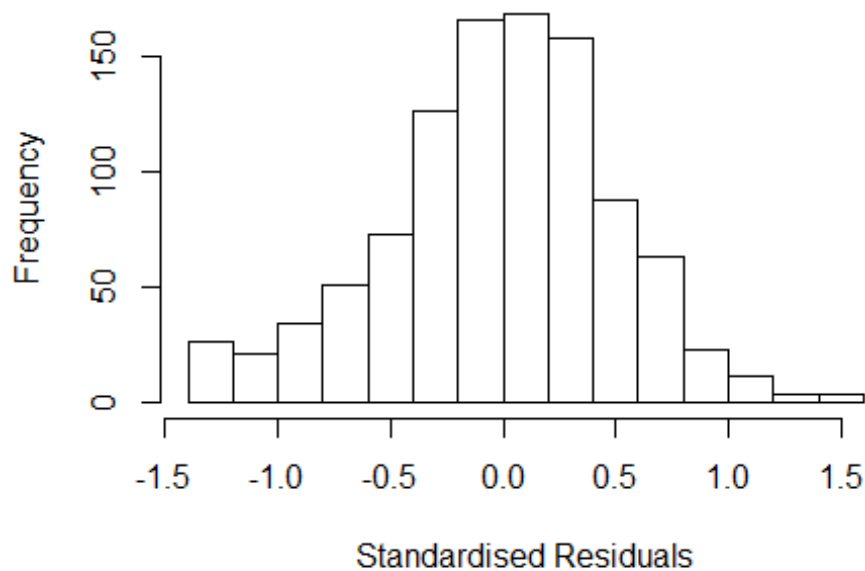
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39226 -0.30122 0.00638 0.29588 1.50005
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8237 0.1500 5.49 5.1e-08 ***
## FirstAuthorFemale1 0.0187 0.0363 0.51 0.60784
## LastAuthorFemale1 -0.0615 0.0396 -1.55 0.12093
## Year1997 0.5334 0.1876 2.84 0.00456 **
## Year1998 0.2997 0.1796 1.67 0.09545 .
## Year1999 0.3490 0.1661 2.10 0.03589 *
## Year2000 0.3410 0.1657 2.06 0.03987 *
## Year2001 0.3794 0.1880 2.02 0.04388 *
## Year2002 0.3918 0.1802 2.17 0.02991 *
## Year2003 0.3687 0.1635 2.26 0.02432 *
## Year2004 0.4610 0.1689 2.73 0.00646 **
## Year2005 0.3942 0.1694 2.33 0.02018 *
```

```

## Year2006          0.5560      0.1648      3.37  0.00077 ***
## Year2007          0.5272      0.1579      3.34  0.00087 ***
## Year2008          0.5253      0.1637      3.21  0.00138 **
## Year2009          0.5406      0.1584      3.41  0.00067 ***
## Year2010          0.5499      0.1578      3.49  0.00051 ***
## Year2011          0.5646      0.1578      3.58  0.00036 ***
## Year2012          0.4396      0.1587      2.77  0.00571 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0517, Adjusted R-squared:  0.0345
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 931 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.254  0.850   0.952   0.889   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.085 1          1.042
## Year              1.085 16          1.003

```

## Residuals from first author



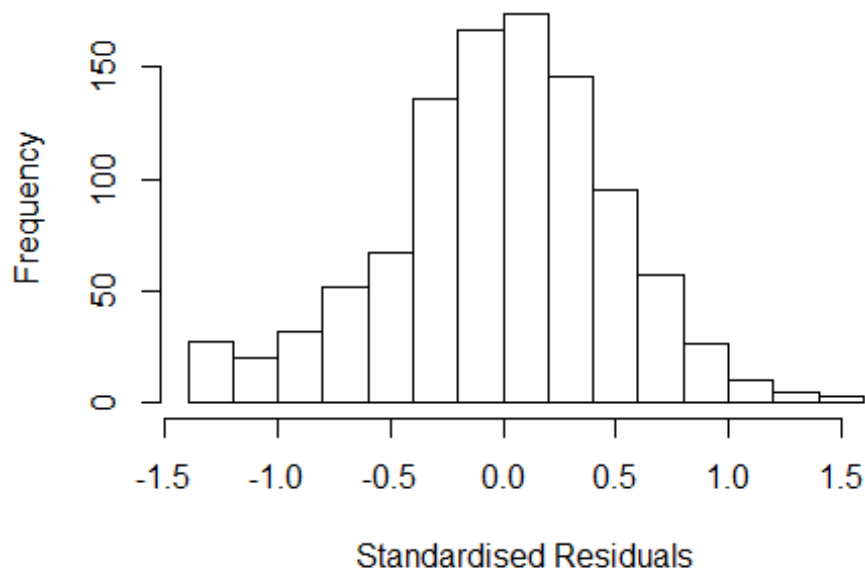
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37644 -0.29602 0.00948 0.29684 1.50778
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.807097 0.148020 5.45 6.3e-08 ***
## FirstAuthorFemale1 0.000505 0.036578 0.01 0.98900
## Year1997 0.539806 0.186135 2.90 0.00381 **
## Year1998 0.304719 0.177645 1.72 0.08660 .
## Year1999 0.360803 0.164565 2.19 0.02858 *
## Year2000 0.344601 0.164098 2.10 0.03598 *
## Year2001 0.376687 0.186676 2.02 0.04387 *
## Year2002 0.407464 0.179458 2.27 0.02339 *
## Year2003 0.374245 0.161699 2.31 0.02085 *
## Year2004 0.470768 0.167481 2.81 0.00504 **
## Year2005 0.406422 0.167764 2.42 0.01559 *
## Year2006 0.567728 0.163047 3.48 0.00052 ***
```

```

## Year2007      0.537969    0.156106    3.45  0.00059 ***
## Year2008      0.534126    0.161990    3.30  0.00101 **
## Year2009      0.546648    0.156526    3.49  0.00050 ***
## Year2010      0.560093    0.155969    3.59  0.00035 ***
## Year2011      0.569341    0.155786    3.65  0.00027 ***
## Year2012      0.448671    0.156684    2.86  0.00428 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.0493, Adjusted R-squared:  0.0331
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 930 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.255  0.849   0.953   0.890   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.1 1      1.049
## Year              1.1 16      1.003

```

## Residuals from last author



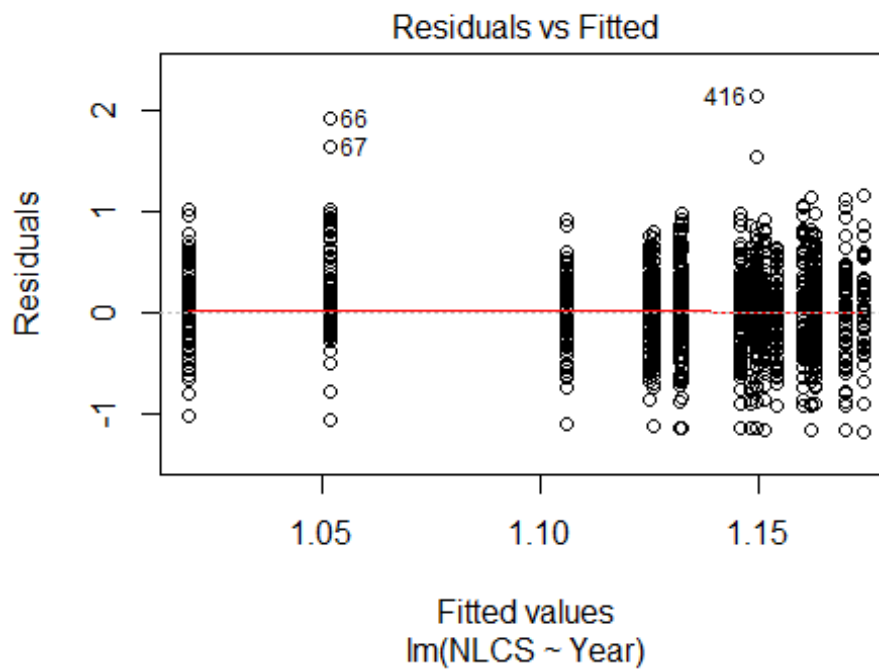
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39213 -0.30029  0.00585  0.29310  1.49728
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.8246     0.1494   5.52 4.4e-08 ***
## LastAuthorFemale1 -0.0552     0.0395  -1.40  0.16239
## Year1997          0.5349     0.1871   2.86  0.00435 **
## Year1998          0.3016     0.1787   1.69  0.09185 .
## Year1999          0.3507     0.1653   2.12  0.03417 *
## Year2000          0.3414     0.1651   2.07  0.03891 *
## Year2001          0.3795     0.1876   2.02  0.04330 *
## Year2002          0.3942     0.1796   2.19  0.02845 *
## Year2003          0.3703     0.1626   2.28  0.02299 *
## Year2004          0.4640     0.1676   2.77  0.00573 **
## Year2005          0.3965     0.1688   2.35  0.01902 *
## Year2006          0.5590     0.1637   3.41  0.00066 ***
```

```

## Year2007          0.5283      0.1572      3.36  0.00081 ***
## Year2008          0.5271      0.1629      3.23  0.00126 **
## Year2009          0.5438      0.1574      3.45  0.00057 ***
## Year2010          0.5531      0.1569      3.53  0.00044 ***
## Year2011          0.5675      0.1566      3.62  0.00031 ***
## Year2012          0.4428      0.1576      2.81  0.00505 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.455
## Multiple R-squared:  0.0516, Adjusted R-squared:  0.0354
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 928 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.256  0.848  0.953  0.888  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1016"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2307"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 223 235 218 199 191 182 183 152 150 140 142 162 168 173 173
## 2011 2012
## 159 133
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 96 74 76 69 48 78 79 78 86 75 90 94 99 107
## 2011 2012

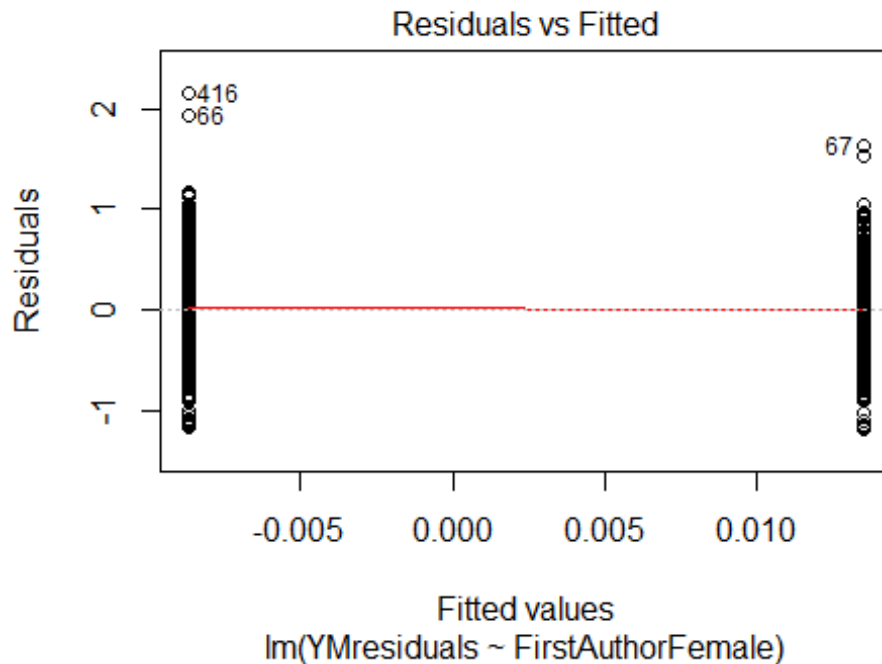
```

```
##      88      88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    84   87   68   67   62   48   69   72   68   74   66   73   86   91   97
## 2011 2012
##    79   73
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 73, df = 16, p-value = 4e-09
```



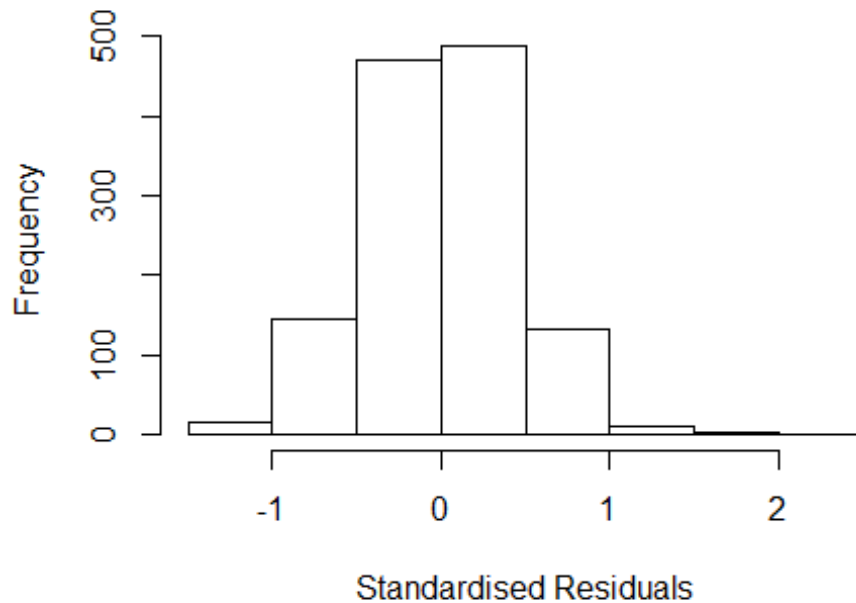
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.3, df = 1, p-value = 0.007
```





```
## [1] "Female first author team size 2018 geometric mean: 4.2241228384697"
## [1] "Male first author team size 2018 geometric mean: 3.63409766218959"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.15229092525843"
## [1] "Male last author team size 2018 geometric mean: 3.75565126038422"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1          1.033
## LastAuthorFemale  1.085 1          1.042
## UniqueAuthors    1.310 4          1.034
## Year             1.423 16          1.011
```

## Residuals from first and last author and team size



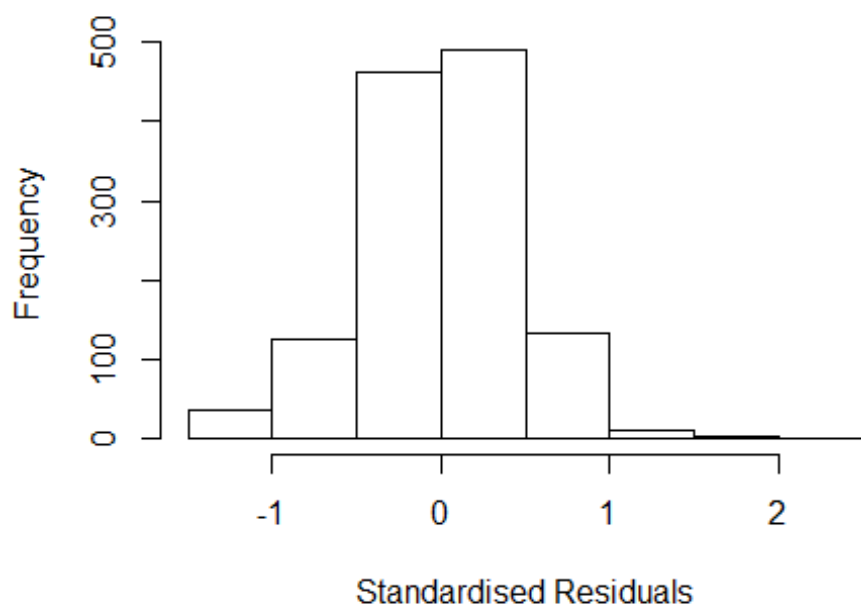
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22954 -0.26182 0.00373 0.26999 2.14539
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.84397 0.09171 9.20 < 2e-16 ***
## FirstAuthorFemale1 0.00516 0.02479 0.21 0.83510
## LastAuthorFemale1 -0.05835 0.02811 -2.08 0.03808 *
## UniqueAuthors2 0.22246 0.06064 3.67 0.00025 ***
## UniqueAuthors3 0.21541 0.05794 3.72 0.00021 ***
## UniqueAuthors4 0.23631 0.06079 3.89 0.00011 ***
## UniqueAuthors5 0.30597 0.05711 5.36 1e-07 ***
## Year1997 0.08418 0.08339 1.01 0.31294
## Year1998 0.11501 0.09588 1.20 0.23058
## Year1999 0.08151 0.08947 0.91 0.36247
```

```

## Year2000      0.13057      0.08935      1.46  0.14418
## Year2001      0.14491      0.09866      1.47  0.14214
## Year2002      0.16500      0.09301      1.77  0.07630 .
## Year2003     -0.04577      0.08558     -0.53  0.59289
## Year2004      0.04937      0.08863      0.56  0.57760
## Year2005      0.10553      0.08102      1.30  0.19298
## Year2006      0.11780      0.08162      1.44  0.14918
## Year2007      0.07591      0.08666      0.88  0.38124
## Year2008      0.05942      0.07749      0.77  0.44336
## Year2009      0.06605      0.07819      0.84  0.39844
## Year2010      0.10422      0.07818      1.33  0.18274
## Year2011      0.06689      0.07872      0.85  0.39565
## Year2012      0.11670      0.08522      1.37  0.17114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.0581, Adjusted R-squared:  0.0414
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,127) are outliers with |weight| = 0 ( < 7.9e-05);
## 110 weights are ~= 1. The remaining 1152 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.107  0.857  0.953  0.896  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           7.91e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample   max.it   best.r.s   k.fast.s   k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev   mts   compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1 1.031
## LastAuthorFemale 1.067 1 1.033
## Year 1.132 16 1.004

```

## Residuals from first and last author

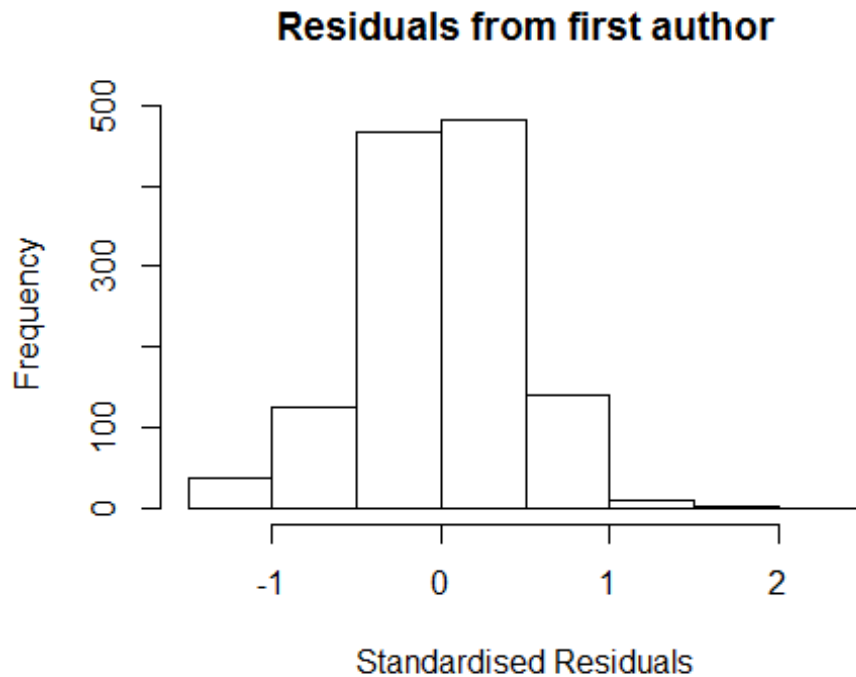


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.24327 -0.26950 0.00441 0.27870 2.16919
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03442 0.07708 13.42 <2e-16 ***
## FirstAuthorFemale1 0.01567 0.02518 0.62 0.534
## LastAuthorFemale1 -0.07019 0.02828 -2.48 0.013 *
## Year1997 0.09239 0.08855 1.04 0.297
## Year1998 0.13848 0.10023 1.38 0.167
## Year1999 0.11429 0.09551 1.20 0.232
## Year2000 0.15539 0.09319 1.67 0.096 .
## Year2001 0.16770 0.10290 1.63 0.103
## Year2002 0.19317 0.10089 1.91 0.056 .
## Year2003 0.00413 0.09183 0.04 0.964
## Year2004 0.09856 0.09413 1.05 0.295
## Year2005 0.15443 0.08699 1.78 0.076 .
```

```

## Year2006          0.14704      0.08774      1.68      0.094 .
## Year2007          0.12430      0.09254      1.34      0.179
## Year2008          0.10140      0.08445      1.20      0.230
## Year2009          0.12319      0.08569      1.44      0.151
## Year2010          0.12892      0.08506      1.52      0.130
## Year2011          0.11490      0.08535      1.35      0.178
## Year2012          0.13741      0.09154      1.50      0.134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0178, Adjusted R-squared:  0.00361
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,127) are outliers with |weight| = 0 ( < 7.9e-05);
## 113 weights are ~ = 1. The remaining 1149 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0661 0.8570 0.9520 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.91e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.06 1          1.030
## Year              1.06 16          1.002

```

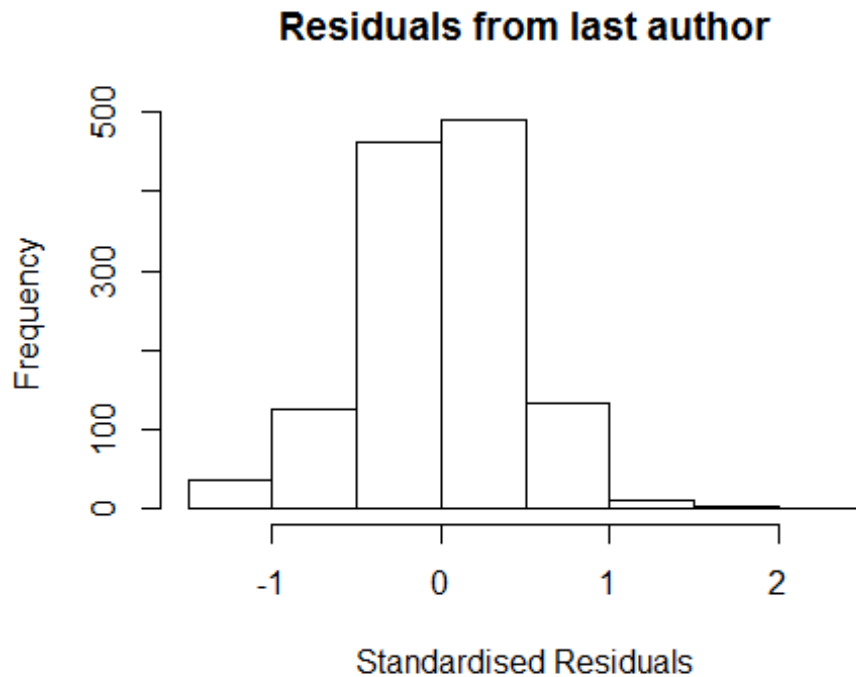


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22446 -0.26933 0.00739 0.28561 2.18236
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01858 0.07709 13.21 <2e-16 ***
## FirstAuthorFemale1 0.01089 0.02527 0.43 0.667
## Year1997 0.09506 0.08921 1.07 0.287
## Year1998 0.13763 0.10013 1.37 0.170
## Year1999 0.11672 0.09585 1.22 0.224
## Year2000 0.15326 0.09327 1.64 0.101
## Year2001 0.16644 0.10342 1.61 0.108
## Year2002 0.19499 0.10170 1.92 0.055 .
## Year2003 -0.00359 0.09269 -0.04 0.969
## Year2004 0.10317 0.09470 1.09 0.276
## Year2005 0.16190 0.08717 1.86 0.064 .
## Year2006 0.15308 0.08797 1.74 0.082 .
```

```

## Year2007          0.12930    0.09344    1.38    0.167
## Year2008          0.10295    0.08501    1.21    0.226
## Year2009          0.12971    0.08605    1.51    0.132
## Year2010          0.12383    0.08604    1.44    0.150
## Year2011          0.12175    0.08601    1.42    0.157
## Year2012          0.13916    0.09272    1.50    0.134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.414
## Multiple R-squared:  0.0135, Adjusted R-squared:  5.17e-05
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,127) are outliers with |weight| = 0 ( < 7.9e-05);
## 122 weights are ~= 1. The remaining 1140 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.063  0.858  0.949  0.895  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.066 1          1.033
## Year            1.066 16          1.002

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23313 -0.26660 0.00677 0.28082 2.16477
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03882 0.07636 13.60 <2e-16 ***
## LastAuthorFemale1 -0.06866 0.02833 -2.42 0.016 *
## Year1997 0.09241 0.08855 1.04 0.297
## Year1998 0.14063 0.10026 1.40 0.161
## Year1999 0.11428 0.09571 1.19 0.233
## Year2000 0.15415 0.09309 1.66 0.098 .
## Year2001 0.17001 0.10278 1.65 0.098 .
## Year2002 0.19431 0.10122 1.92 0.055 .
## Year2003 0.00525 0.09204 0.06 0.954
## Year2004 0.10026 0.09429 1.06 0.288
## Year2005 0.15716 0.08706 1.81 0.071 .
## Year2006 0.14992 0.08810 1.70 0.089 .
```

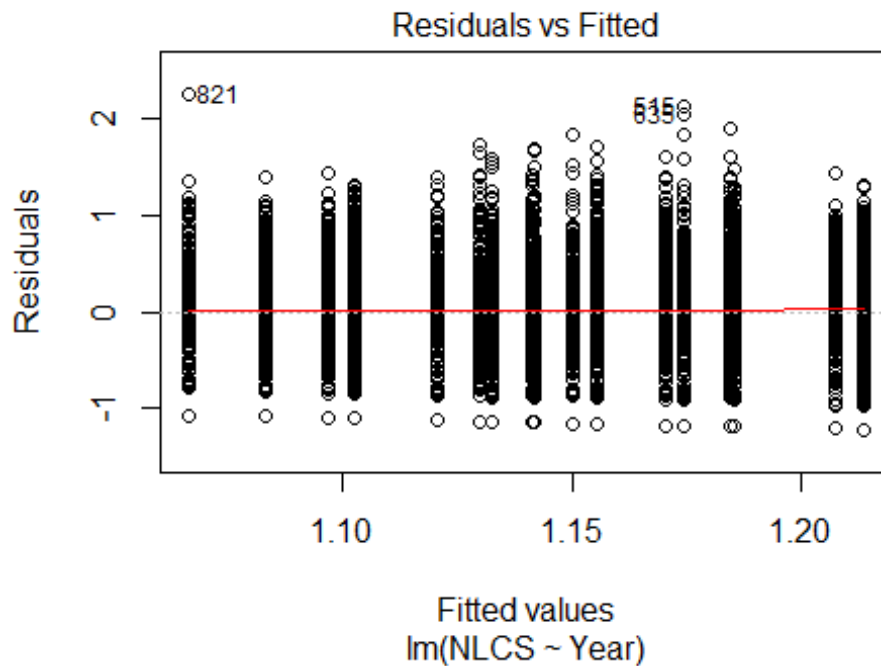


```

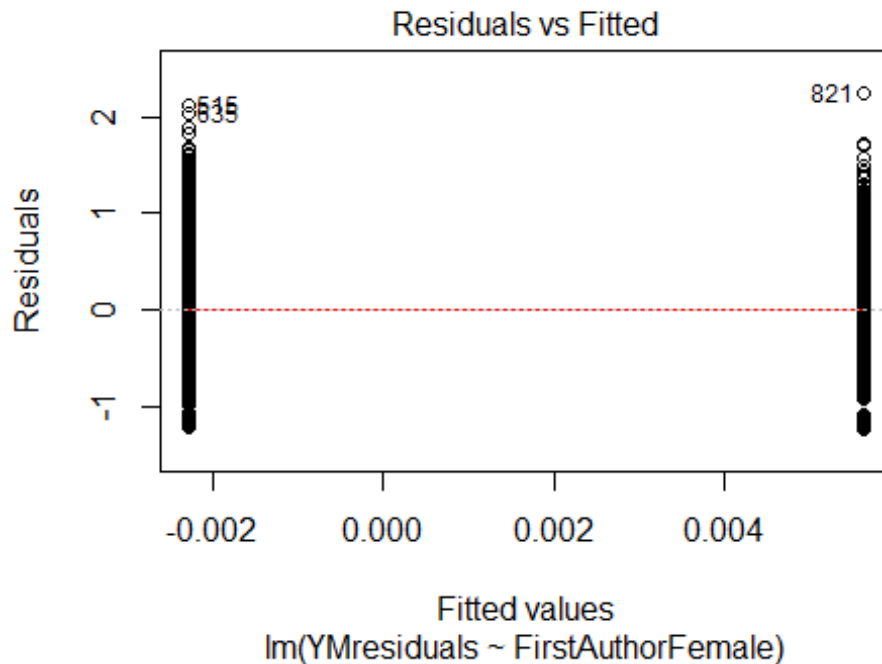
## Year2007      0.12669      0.09295      1.36      0.173
## Year2008      0.10441      0.08459      1.23      0.217
## Year2009      0.12553      0.08587      1.46      0.144
## Year2010      0.13018      0.08530      1.53      0.127
## Year2011      0.11830      0.08562      1.38      0.167
## Year2012      0.13983      0.09170      1.52      0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0175, Adjusted R-squared:  0.00413
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(22,127) are outliers with |weight| = 0 ( < 7.9e-05);
## 116 weights are ~= 1. The remaining 1146 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0602 0.8560 0.9530 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1264"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2308"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 381 362 414 494 492 475 467 457 410 423 449 555 563 575 589
## 2011 2012
## 683 721
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 207 209 208 242 230 180 324 323 291 323 340 436 442 447 466

```

```
## 2011 2012
## 558 569
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 204 203 235 222 173 305 310 279 310 327 406 422 431 438
## 2011 2012
## 522 526
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

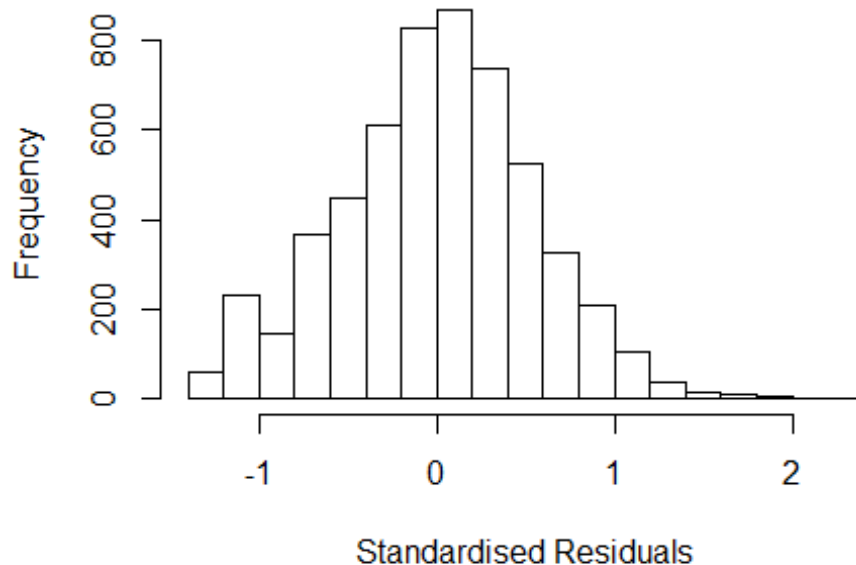


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.49, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 2.09100895816843"
## [1] "Male first author team size 2018 geometric mean: 2.00523680285964"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.91863809358224"
## [1] "Male last author team size 2018 geometric mean: 2.09958663314421"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 39000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.277 1      1.130
## LastAuthorFemale  1.266 1      1.125
## UniqueAuthors    1.065 4      1.008
## Year              1.087 16     1.003
```

## Residuals from first and last author and team size



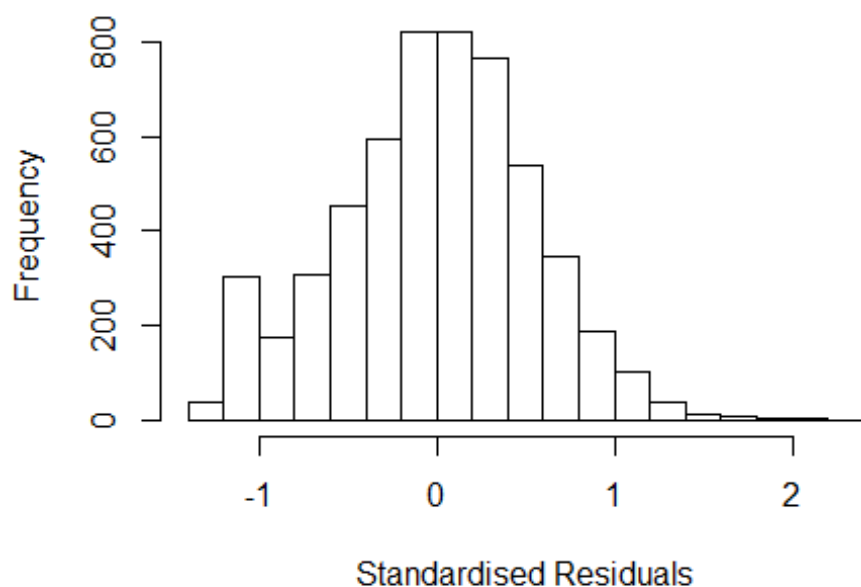
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3855 -0.3506 0.0167 0.3514 2.3252
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.033208 0.048628 21.25 < 2e-16 ***
## FirstAuthorFemale1 0.007199 0.018647 0.39 0.70
## LastAuthorFemale1 -0.015741 0.019404 -0.81 0.42
## UniqueAuthors2 0.200791 0.017662 11.37 < 2e-16 ***
## UniqueAuthors3 0.258492 0.023040 11.22 < 2e-16 ***
## UniqueAuthors4 0.232928 0.030840 7.55 5.0e-14 ***
## UniqueAuthors5 0.242832 0.033156 7.32 2.8e-13 ***
## Year1997 0.022763 0.063557 0.36 0.72
## Year1998 -0.037848 0.063962 -0.59 0.55
## Year1999 0.015509 0.060368 0.26 0.80
```

```

## Year2000      0.000303  0.059503  0.01  1.00
## Year2001      0.005820  0.065954  0.09  0.93
## Year2002      0.009415  0.057311  0.16  0.87
## Year2003      0.068518  0.056204  1.22  0.22
## Year2004      0.041274  0.058312  0.71  0.48
## Year2005      0.022013  0.057700  0.38  0.70
## Year2006     -0.030100  0.057187 -0.53  0.60
## Year2007     -0.033369  0.055449 -0.60  0.55
## Year2008      0.086554  0.054620  1.58  0.11
## Year2009      0.035154  0.055258  0.64  0.52
## Year2010     -0.036463  0.054745 -0.67  0.51
## Year2011      0.004366  0.054038  0.08  0.94
## Year2012      0.032416  0.054705  0.59  0.55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0464, Adjusted R-squared:  0.0426
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 471 weights are ~= 1. The remaining 5042 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.011  0.859  0.950  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.299 1      1.140
## LastAuthorFemale  1.289 1      1.135
## Year              1.027 16      1.001

```

## Residuals from first and last author



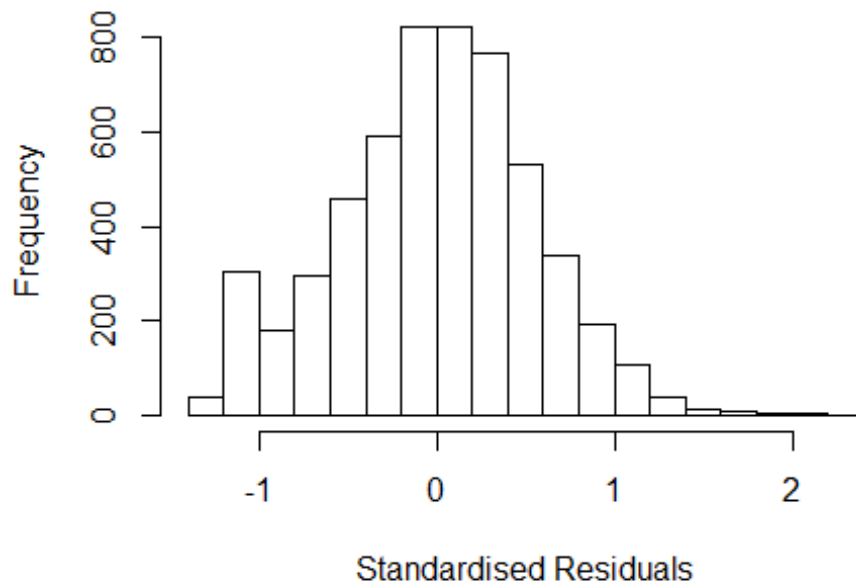
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2595 -0.3543 0.0171 0.3544 2.2491
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11066 0.04980 22.30 <2e-16 ***
## FirstAuthorFemale1 0.02561 0.01911 1.34 0.180
## LastAuthorFemale1 -0.03042 0.01996 -1.52 0.128
## Year1997 0.04946 0.06495 0.76 0.446
## Year1998 -0.04295 0.06627 -0.65 0.517
## Year1999 0.02749 0.06238 0.44 0.659
## Year2000 0.01762 0.06136 0.29 0.774
## Year2001 0.00885 0.06752 0.13 0.896
## Year2002 0.02127 0.05910 0.36 0.719
## Year2003 0.10736 0.05785 1.86 0.064 .
## Year2004 0.06812 0.05988 1.14 0.255
## Year2005 0.03812 0.05921 0.64 0.520
```

```

## Year2006      -0.01054    0.05921   -0.18    0.859
## Year2007      -0.00829    0.05686   -0.15    0.884
## Year2008       0.12328    0.05638    2.19    0.029 *
## Year2009       0.06719    0.05690    1.18    0.238
## Year2010      -0.00679    0.05661   -0.12    0.905
## Year2011       0.04235    0.05565    0.76    0.447
## Year2012       0.06995    0.05611    1.25    0.213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.00661,    Adjusted R-squared:  0.00335
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 467 weights are ~= 1. The remaining 5046 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0334 0.8610 0.9490 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1      1.009
## Year      1.017 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.241 -0.358 0.017 0.354 2.239
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10800 0.04973 22.28 <2e-16 ***
## FirstAuthorFemale1 0.00911 0.01707 0.53 0.594
## Year1997 0.04982 0.06499 0.77 0.443
## Year1998 -0.04393 0.06627 -0.66 0.507
## Year1999 0.02663 0.06239 0.43 0.670
## Year2000 0.01796 0.06143 0.29 0.770
## Year2001 0.00789 0.06751 0.12 0.907
## Year2002 0.02034 0.05906 0.34 0.731
## Year2003 0.10635 0.05781 1.84 0.066 .
## Year2004 0.06755 0.05987 1.13 0.259
## Year2005 0.03878 0.05920 0.66 0.512
## Year2006 -0.01135 0.05918 -0.19 0.848
```

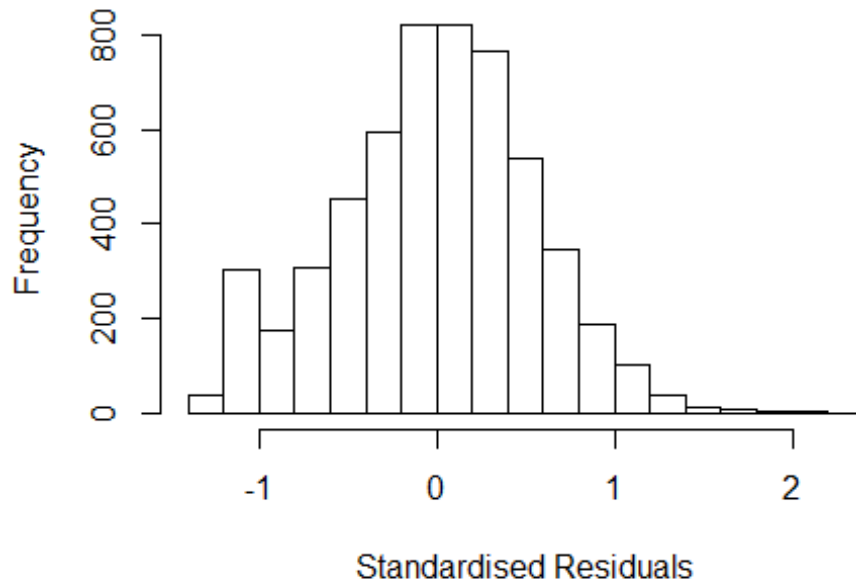


```

## Year2007          -0.00963    0.05683   -0.17    0.865
## Year2008          0.12357    0.05633    2.19    0.028 *
## Year2009          0.06609    0.05690    1.16    0.245
## Year2010         -0.00715    0.05664   -0.13    0.900
## Year2011          0.04101    0.05565    0.74    0.461
## Year2012          0.06982    0.05611    1.24    0.213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.00623,    Adjusted R-squared:  0.00316
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 462 weights are ~= 1. The remaining 5051 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0361 0.8620 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.01 1      1.005
## Year              1.01 16      1.000

```

## Residuals from last author



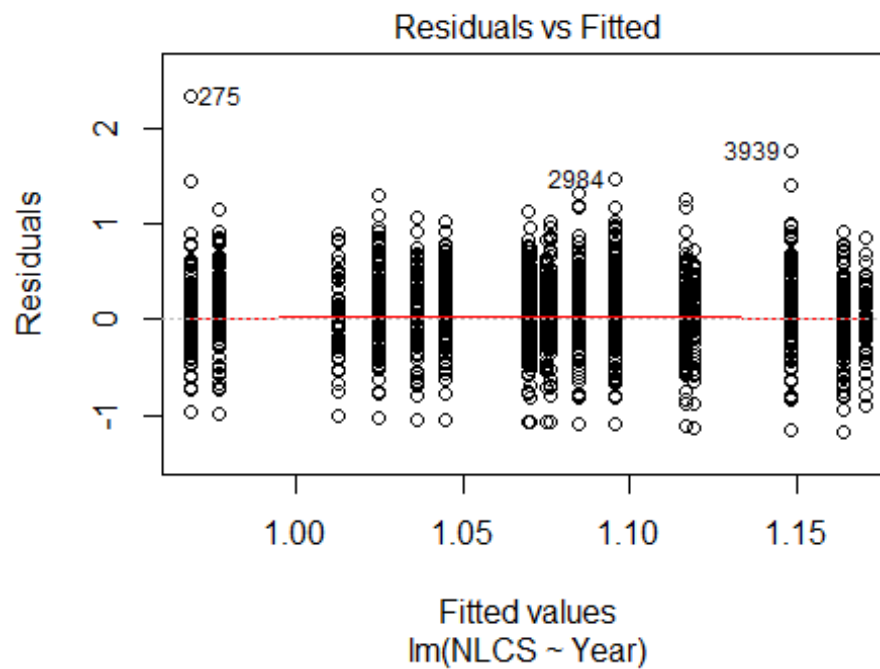
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2379 -0.3562 0.0182 0.3578 2.2582
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11266 0.04976 22.36 <2e-16 ***
## LastAuthorFemale1 -0.01575 0.01771 -0.89 0.374
## Year1997 0.05014 0.06487 0.77 0.440
## Year1998 -0.04309 0.06622 -0.65 0.515
## Year1999 0.02761 0.06232 0.44 0.658
## Year2000 0.01955 0.06133 0.32 0.750
## Year2001 0.01009 0.06747 0.15 0.881
## Year2002 0.02176 0.05904 0.37 0.713
## Year2003 0.10837 0.05780 1.87 0.061 .
## Year2004 0.07053 0.05979 1.18 0.238
## Year2005 0.04080 0.05902 0.69 0.489
## Year2006 -0.00910 0.05910 -0.15 0.878
```

```

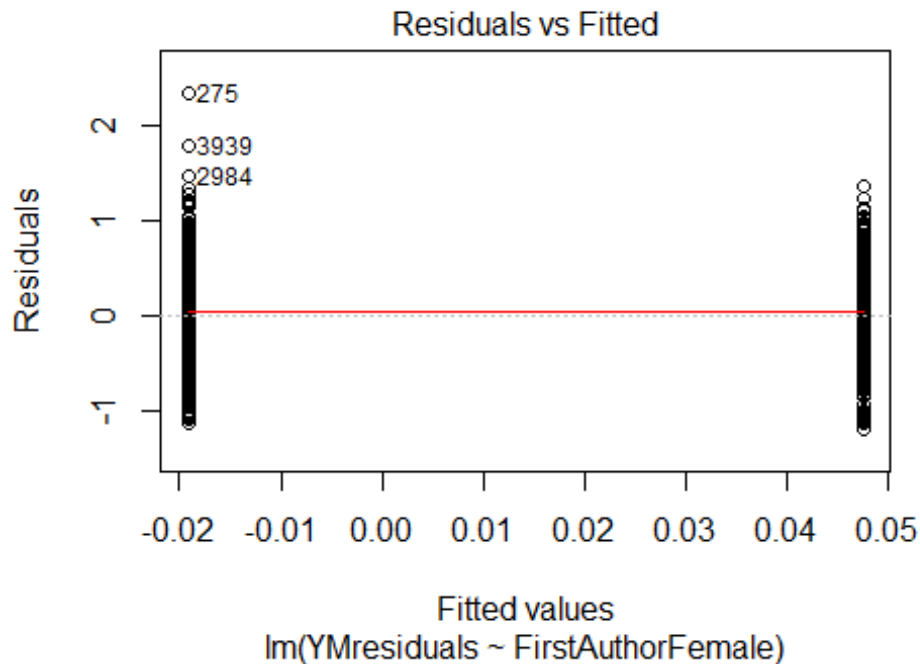
## Year2007          -0.00631      0.05675    -0.11      0.911
## Year2008           0.12524      0.05627      2.23      0.026 *
## Year2009           0.06896      0.05683      1.21      0.225
## Year2010          -0.00475      0.05655    -0.08      0.933
## Year2011           0.04363      0.05556      0.79      0.432
## Year2012           0.07191      0.05603      1.28      0.199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.00634,    Adjusted R-squared:  0.00327
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 477 weights are ~= 1. The remaining 5036 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0309 0.8600 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.81e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5513"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2309"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 144 175 170 161 197 140 148 164 144 146 157 183 180 223 234
## 2011 2012
## 207 214
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 91 81 89 101 58 102 109 101 98 118 145 139 168 171
## 2011 2012

```

```
## 147 161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 84 77 84 95 55 95 101 97 87 107 130 130 160 156
## 2011 2012
## 139 145
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 54, df = 16, p-value = 6e-06
```

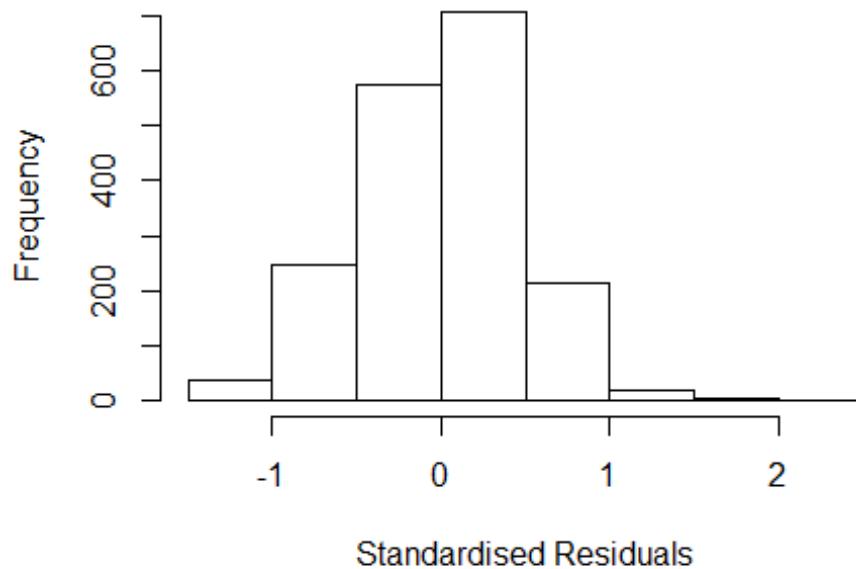


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.48, df = 1, p-value = 0.5
```



```
## [1] "Female first author team size 2018 geometric mean: 2.83841569160087"
## [1] "Male first author team size 2018 geometric mean: 2.59006056767764"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.84127016550372"
## [1] "Male last author team size 2018 geometric mean: 2.61023186239584"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.174 1      1.084
## LastAuthorFemale  1.104 1      1.051
## UniqueAuthors    1.333 4      1.037
## Year              1.592 16     1.015
```

## Residuals from first and last author and team size



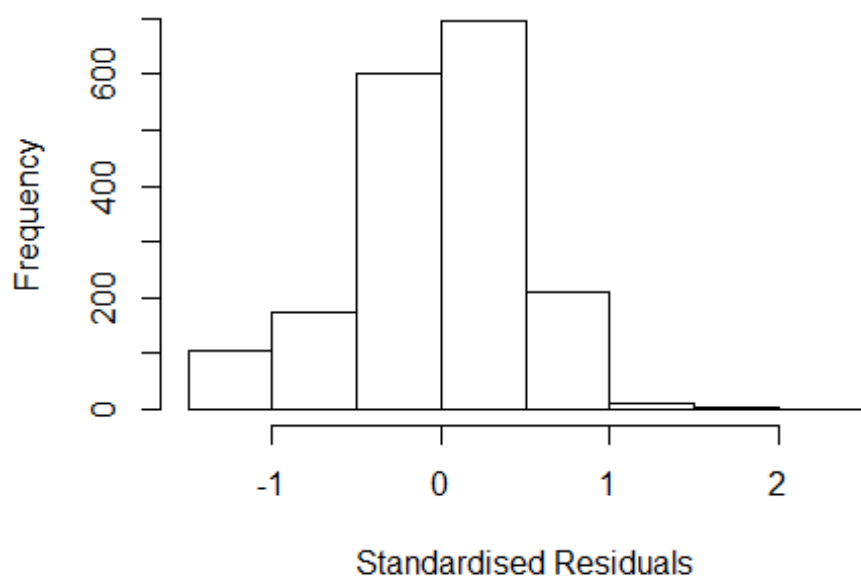
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2043 -0.2834 0.0231 0.3099 2.3291
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.91489 0.06269 14.59 < 2e-16 ***
## FirstAuthorFemale1 0.07110 0.02571 2.76 0.0058 **
## LastAuthorFemale1 -0.01458 0.03099 -0.47 0.6381
## UniqueAuthors2 0.18543 0.03567 5.20 2.3e-07 ***
## UniqueAuthors3 0.19625 0.03627 5.41 7.1e-08 ***
## UniqueAuthors4 0.24060 0.03961 6.07 1.5e-09 ***
## UniqueAuthors5 0.28627 0.04009 7.14 1.4e-12 ***
## Year1997 -0.13346 0.07779 -1.72 0.0864 .
## Year1998 0.07744 0.07589 1.02 0.3077
## Year1999 0.02205 0.07345 0.30 0.7640
```

```

## Year2000      0.10406      0.07331      1.42      0.1559
## Year2001      0.09288      0.07800      1.19      0.2339
## Year2002      0.01928      0.07879      0.24      0.8067
## Year2003      0.01347      0.07328      0.18      0.8541
## Year2004     -0.03100      0.07808     -0.40      0.6914
## Year2005     -0.00444      0.08186     -0.05      0.9567
## Year2006     -0.08674      0.08259     -1.05      0.2938
## Year2007     -0.05422      0.08199     -0.66      0.5085
## Year2008     -0.02386      0.07823     -0.31      0.7604
## Year2009     -0.00174      0.07455     -0.02      0.9814
## Year2010     -0.02434      0.07378     -0.33      0.7415
## Year2011     -0.01141      0.07600     -0.15      0.8806
## Year2012      0.01656      0.07847      0.21      0.8329
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0663, Adjusted R-squared:  0.0547
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 109 is an outlier with |weight| = 0 ( < 5.6e-05);
## 160 weights are ~= 1. The remaining 1638 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.101  0.853  0.950  0.896  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.56e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.198 1 1.095
## LastAuthorFemale 1.097 1 1.047
## Year 1.250 16 1.007

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2149 -0.2910 0.0107 0.3077 2.3897
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0025 0.0575 17.42 < 2e-16 ***
## FirstAuthorFemale1 0.0893 0.0263 3.39 0.00071 ***
## LastAuthorFemale1 -0.0399 0.0314 -1.27 0.20356
## Year1997 -0.0963 0.0774 -1.24 0.21335
## Year1998 0.1270 0.0712 1.78 0.07452 .
## Year1999 0.0727 0.0706 1.03 0.30296
## Year2000 0.1647 0.0703 2.34 0.01933 *
## Year2001 0.1616 0.0755 2.14 0.03249 *
## Year2002 0.0839 0.0748 1.12 0.26219
## Year2003 0.0829 0.0712 1.16 0.24419
## Year2004 0.0201 0.0760 0.26 0.79167
## Year2005 0.0779 0.0819 0.95 0.34106
```

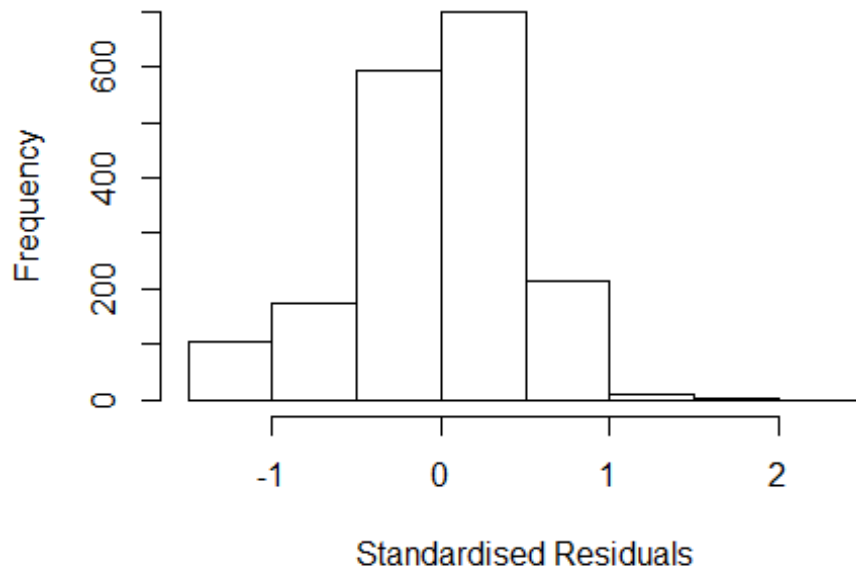


```

## Year2006          -0.0321      0.0828    -0.39  0.69784
## Year2007           0.0308      0.0798      0.39  0.69982
## Year2008           0.0616      0.0742      0.83  0.40687
## Year2009           0.0925      0.0702      1.32  0.18785
## Year2010           0.0715      0.0698      1.03  0.30530
## Year2011           0.0800      0.0723      1.11  0.26845
## Year2012           0.1231      0.0733      1.68  0.09357 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.0223, Adjusted R-squared:  0.0124
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 109 is an outlier with |weight| = 0 ( < 5.6e-05);
## 155 weights are ~= 1. The remaining 1643 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0643 0.8600 0.9480 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.176 1          1.084
## Year              1.176 16          1.005

```

## Residuals from first author



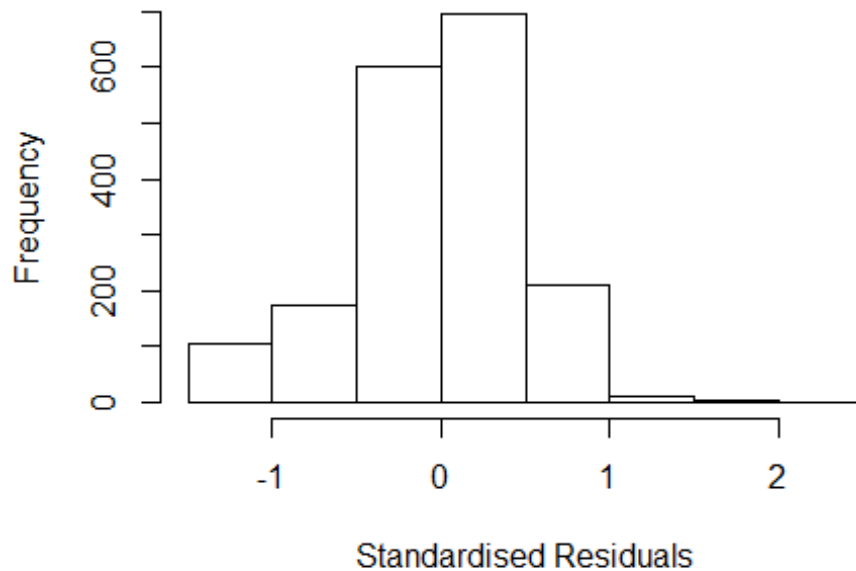
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1996 -0.2929 0.0134 0.3063 2.3919
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9979 0.0577 17.29 <2e-16 ***
## FirstAuthorFemale1 0.0801 0.0263 3.04 0.0024 **
## Year1997 -0.0939 0.0777 -1.21 0.2270
## Year1998 0.1257 0.0714 1.76 0.0786 .
## Year1999 0.0719 0.0707 1.02 0.3095
## Year2000 0.1645 0.0706 2.33 0.0200 *
## Year2001 0.1644 0.0757 2.17 0.0300 *
## Year2002 0.0820 0.0751 1.09 0.2748
## Year2003 0.0835 0.0713 1.17 0.2420
## Year2004 0.0211 0.0761 0.28 0.7817
## Year2005 0.0820 0.0817 1.00 0.3156
## Year2006 -0.0304 0.0831 -0.37 0.7149
```

```

## Year2007          0.0309      0.0797      0.39      0.6986
## Year2008          0.0623      0.0743      0.84      0.4017
## Year2009          0.0918      0.0705      1.30      0.1931
## Year2010          0.0704      0.0700      1.01      0.3143
## Year2011          0.0787      0.0724      1.09      0.2772
## Year2012          0.1215      0.0734      1.66      0.0980 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.0214, Adjusted R-squared:  0.0121
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 109 is an outlier with |weight| = 0 ( < 5.6e-05);
## 160 weights are ~= 1. The remaining 1638 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0616 0.8620 0.9470 0.8920 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.076 1          1.037
## Year          1.076 16          1.002

```

## Residuals from last author



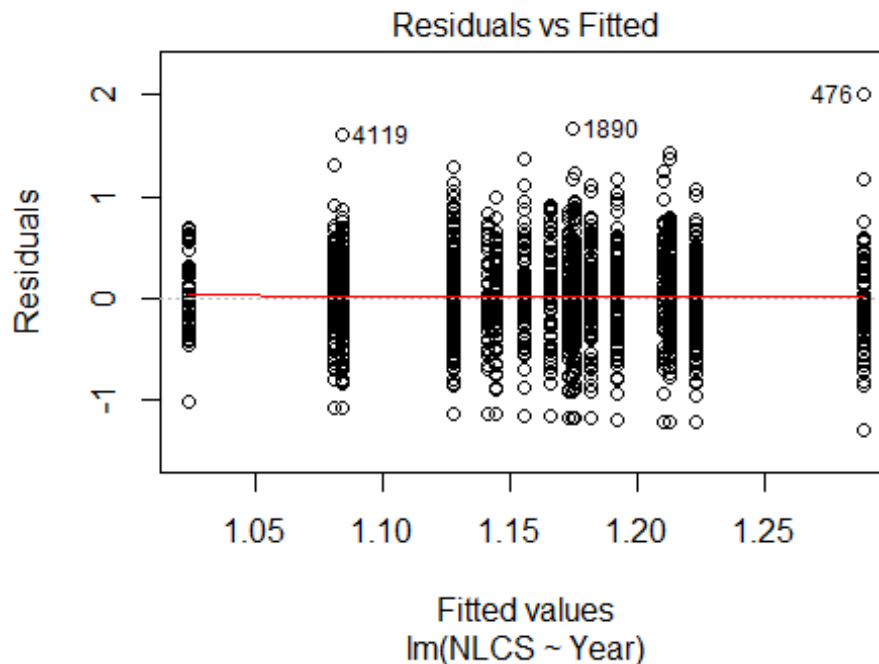
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1494 -0.2840 0.0136 0.3100 2.3809
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0177 0.0576 17.66 <2e-16 ***
## LastAuthorFemale1 -0.0122 0.0309 -0.39 0.694
## Year1997 -0.1026 0.0773 -1.33 0.185
## Year1998 0.1187 0.0713 1.67 0.096 .
## Year1999 0.0733 0.0710 1.03 0.302
## Year2000 0.1631 0.0705 2.31 0.021 *
## Year2001 0.1533 0.0753 2.04 0.042 *
## Year2002 0.0802 0.0753 1.06 0.287
## Year2003 0.0867 0.0714 1.21 0.225
## Year2004 0.0276 0.0762 0.36 0.717
## Year2005 0.0895 0.0821 1.09 0.276
## Year2006 -0.0331 0.0829 -0.40 0.690
```

```

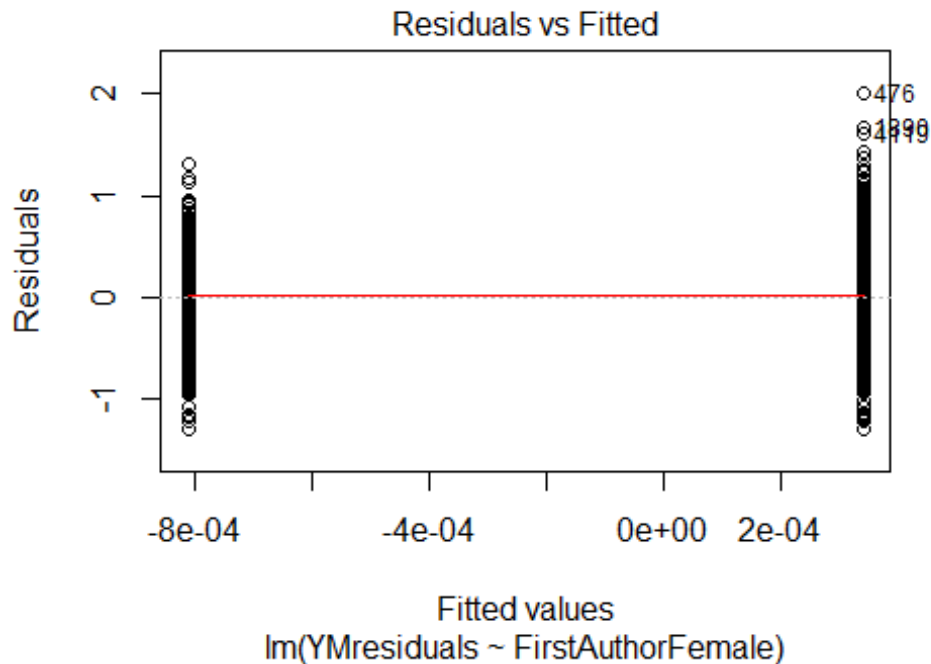
## Year2007          0.0473      0.0793      0.60      0.551
## Year2008          0.0783      0.0739      1.06      0.289
## Year2009          0.1037      0.0703      1.47      0.141
## Year2010          0.0834      0.0698      1.20      0.232
## Year2011          0.0948      0.0721      1.32      0.189
## Year2012          0.1317      0.0734      1.80      0.073 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.0161, Adjusted R-squared:  0.00669
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 109 is an outlier with |weight| = 0 ( < 5.6e-05);
## 156 weights are ~= 1. The remaining 1642 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0741 0.8590 0.9490 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.56e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1799"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2310"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 290 317 373 344 341 283 284 196 238 235 287 288 287 261 308
## 2011 2012
## 309 274
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 82 84 83 86 58 110 64 89 103 134 147 147 144 155

```

```
## 2011 2012
## 167 156
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 74 77 77 81 52 99 60 78 94 123 128 130 133 146
## 2011 2012
## 147 143
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 3e-04
```

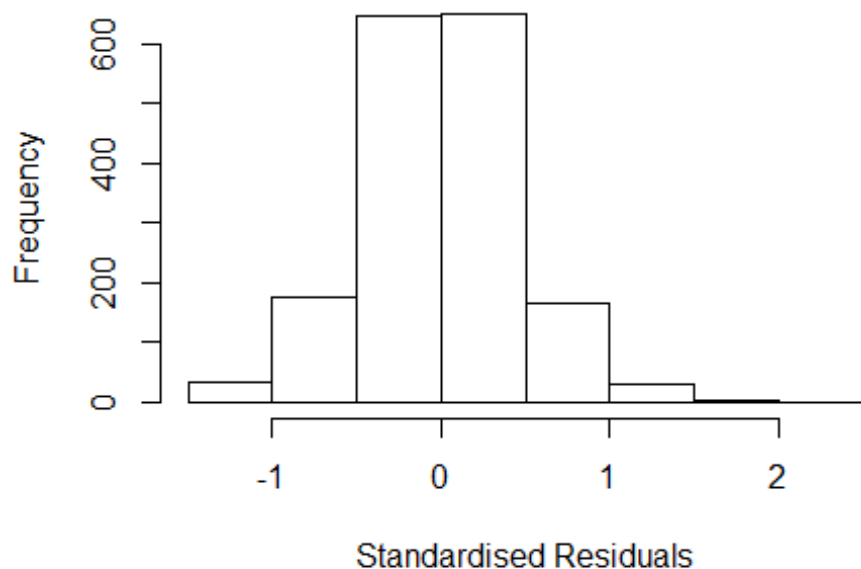


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.3, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 4.08231722293042"
## [1] "Male first author team size 2018 geometric mean: 3.69833380739999"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.95913134837502"
## [1] "Male last author team size 2018 geometric mean: 3.79174410079646"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1      1.023
## LastAuthorFemale  1.063 1      1.031
## UniqueAuthors    1.230 4      1.026
## Year              1.317 16     1.009
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.37253 -0.26681 -0.00049 0.28089 2.02832
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98634 0.08660 11.39 < 2e-16 ***
## FirstAuthorFemale1 -0.00712 0.02326 -0.31 0.760
## LastAuthorFemale1 -0.00759 0.02524 -0.30 0.764
## UniqueAuthors2 0.20871 0.04724 4.42 1.1e-05 ***
## UniqueAuthors3 0.28588 0.04594 6.22 6.1e-10 ***
## UniqueAuthors4 0.28200 0.04756 5.93 3.7e-09 ***
## UniqueAuthors5 0.32600 0.04565 7.14 1.4e-12 ***
## Year1997 0.07263 0.09052 0.80 0.422
## Year1998 -0.00638 0.09869 -0.06 0.948
## Year1999 0.06019 0.09702 0.62 0.535
```

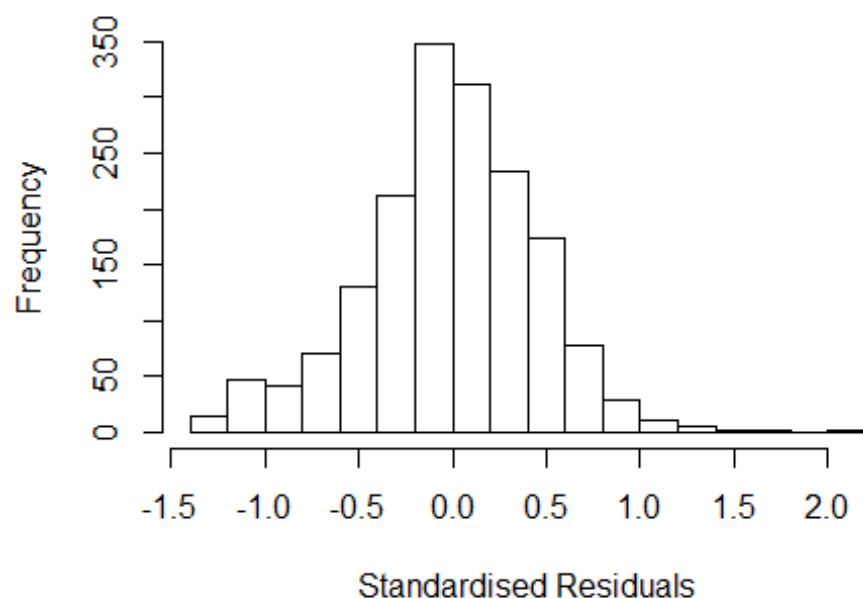


```

## Year2000      -0.05033    0.09113   -0.55    0.581
## Year2001      -0.11350    0.08923   -1.27    0.204
## Year2002       0.06748    0.09036    0.75    0.455
## Year2003      -0.17170    0.09557   -1.80    0.073 .
## Year2004      -0.00385    0.09093   -0.04    0.966
## Year2005      -0.06186    0.08563   -0.72    0.470
## Year2006      -0.04338    0.08357   -0.52    0.604
## Year2007      -0.12024    0.08690   -1.38    0.167
## Year2008      -0.04909    0.08372   -0.59    0.558
## Year2009      -0.13144    0.08254   -1.59    0.112
## Year2010      -0.03534    0.08307   -0.43    0.671
## Year2011      -0.11130    0.08732   -1.27    0.203
## Year2012      -0.06622    0.08262   -0.80    0.423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0693, Adjusted R-squared:  0.0571
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 112 is an outlier with |weight| = 0 ( < 5.9e-05);
## 143 weights are ~= 1. The remaining 1563 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0076 0.8630 0.9500 0.8920 0.9870 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.86e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1 1.019
## LastAuthorFemale 1.054 1 1.027
## Year 1.091 16 1.003

```

## Residuals from first and last author



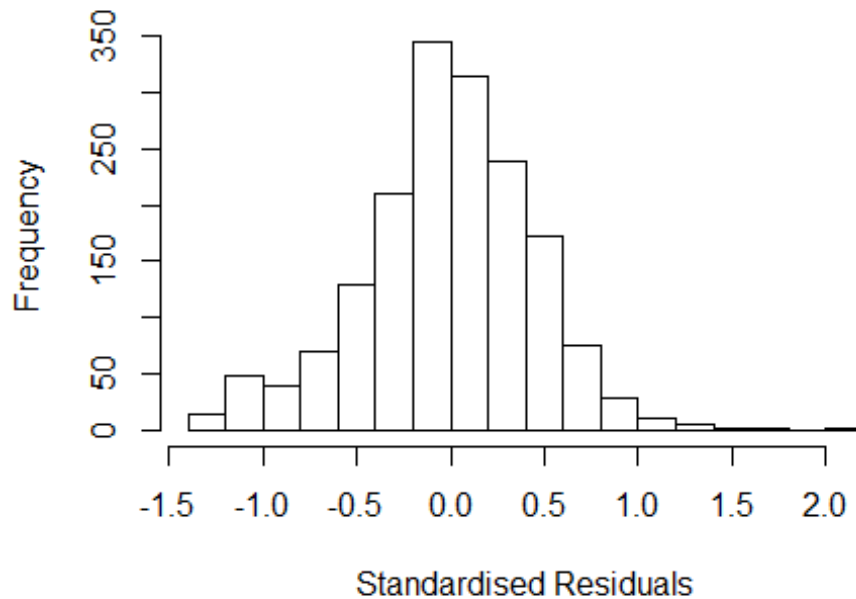
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28860 -0.27707 -0.00328 0.27895 2.01467
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17728 0.07272 16.19 <2e-16 ***
## FirstAuthorFemale1 0.00727 0.02340 0.31 0.76
## LastAuthorFemale1 -0.01183 0.02525 -0.47 0.64
## Year1997 0.10405 0.08873 1.17 0.24
## Year1998 0.01621 0.09481 0.17 0.86
## Year1999 0.06418 0.09284 0.69 0.49
## Year2000 -0.01409 0.08986 -0.16 0.88
## Year2001 -0.07680 0.08902 -0.86 0.39
## Year2002 0.08336 0.08796 0.95 0.34
## Year2003 -0.12188 0.09433 -1.29 0.20
## Year2004 0.05036 0.08558 0.59 0.56
## Year2005 0.00041 0.08267 0.00 1.00
```

```

## Year2006          -0.00491    0.07962   -0.06    0.95
## Year2007          -0.07686    0.08379   -0.92    0.36
## Year2008           0.01595    0.08056    0.20    0.84
## Year2009          -0.06958    0.07984   -0.87    0.38
## Year2010           0.02332    0.07970    0.29    0.77
## Year2011          -0.04373    0.08453   -0.52    0.61
## Year2012          -0.00387    0.07967   -0.05    0.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.00469
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 112 is an outlier with |weight| = 0 ( < 5.9e-05);
## 156 weights are ~= 1. The remaining 1550 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0667 0.8600 0.9500 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.86e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1          1.019
## Year              1.039 16          1.001

```

## Residuals from first author



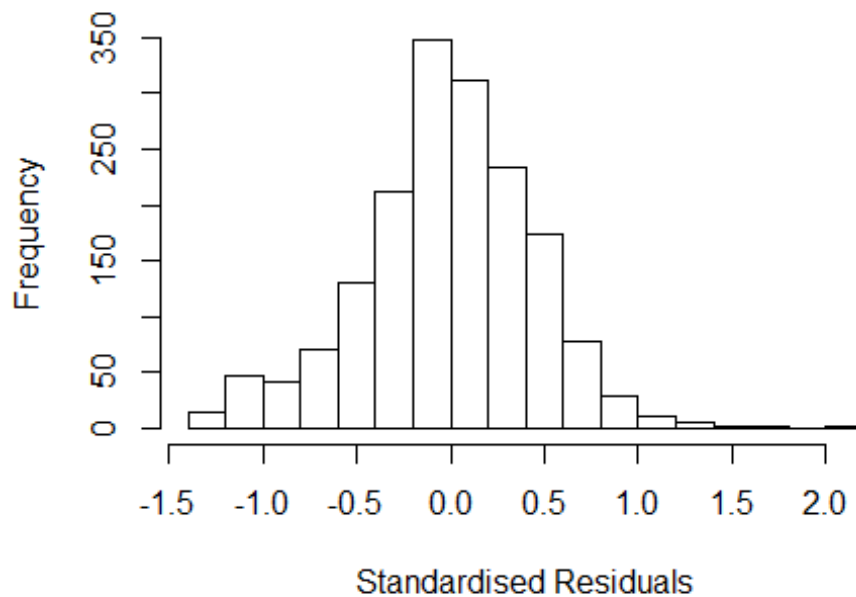
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28599 -0.27706 -0.00143  0.28083  2.01636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17396    0.07204   16.30  <2e-16 ***
## FirstAuthorFemale1 0.00636    0.02341    0.27   0.79
## Year1997        0.10568    0.08844    1.19   0.23
## Year1998        0.01749    0.09450    0.19   0.85
## Year1999        0.06500    0.09282    0.70   0.48
## Year2000       -0.01313    0.08978   -0.15   0.88
## Year2001       -0.07496    0.08881   -0.84   0.40
## Year2002        0.08409    0.08793    0.96   0.34
## Year2003       -0.12157    0.09438   -1.29   0.20
## Year2004        0.05240    0.08525    0.61   0.54
## Year2005        0.00178    0.08257    0.02   0.98
## Year2006       -0.00480    0.07967   -0.06   0.95
```

```

## Year2007      -0.07614    0.08379   -0.91    0.36
## Year2008      0.01676    0.08059    0.21    0.84
## Year2009     -0.06820    0.07974   -0.86    0.39
## Year2010      0.02445    0.07967    0.31    0.76
## Year2011     -0.04152    0.08416   -0.49    0.62
## Year2012     -0.00278    0.07959   -0.03    0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00513
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 112 is an outlier with |weight| = 0 ( < 5.9e-05);
## 154 weights are ~= 1. The remaining 1552 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0739 0.8610 0.9500 0.8920 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.86e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.053 1          1.026
## Year            1.053 16          1.002

```

## Residuals from last author



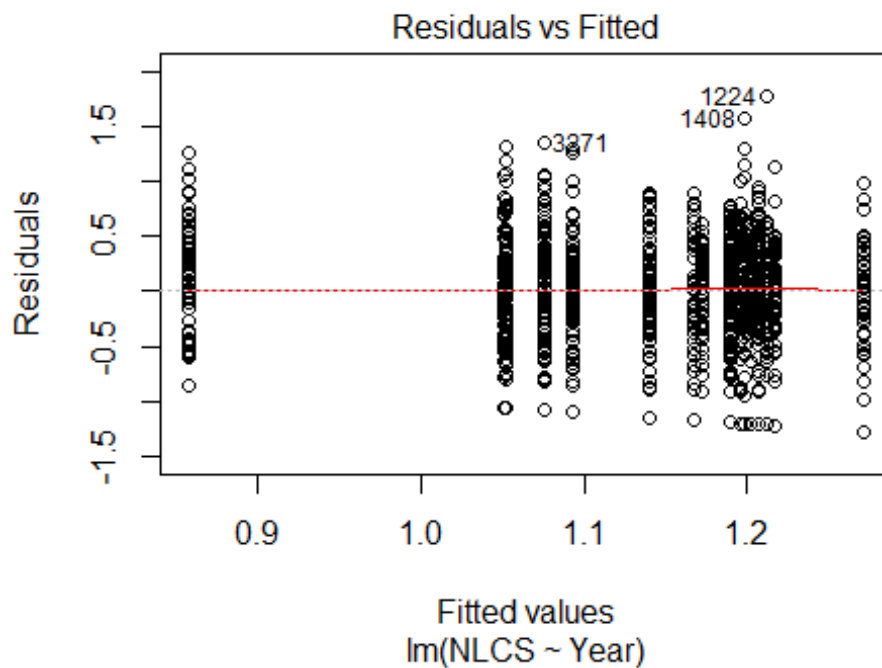
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28308 -0.27834 -0.00369 0.27966 2.01292
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.178839 0.072411 16.28 <2e-16 ***
## LastAuthorFemale1 -0.011109 0.025271 -0.44 0.66
## Year1997 0.104243 0.088660 1.18 0.24
## Year1998 0.016942 0.094745 0.18 0.86
## Year1999 0.064146 0.092765 0.69 0.49
## Year2000 -0.014180 0.089749 -0.16 0.87
## Year2001 -0.075920 0.088829 -0.85 0.39
## Year2002 0.083180 0.087865 0.95 0.34
## Year2003 -0.122066 0.094275 -1.29 0.20
## Year2004 0.050800 0.085455 0.59 0.55
## Year2005 0.000875 0.082548 0.01 0.99
## Year2006 -0.003941 0.079481 -0.05 0.96
```

```

## Year2007      -0.075881    0.083602   -0.91    0.36
## Year2008      0.017200    0.080356    0.21    0.83
## Year2009     -0.069039    0.079811   -0.87    0.39
## Year2010      0.024160    0.079647    0.30    0.76
## Year2011     -0.043254    0.084438   -0.51    0.61
## Year2012     -0.003154    0.079538   -0.04    0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.00522
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 112 is an outlier with |weight| = 0 ( < 5.9e-05);
## 157 weights are ~= 1. The remaining 1549 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0671 0.8580 0.9490 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.86e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1707"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2311"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 157 171 208 176 199 196 150 101 119 118 162 161 192 198 212
## 2011 2012
## 223 205
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 45 64 59 45 53 48 62 36 47 58 78 77 102 91 91

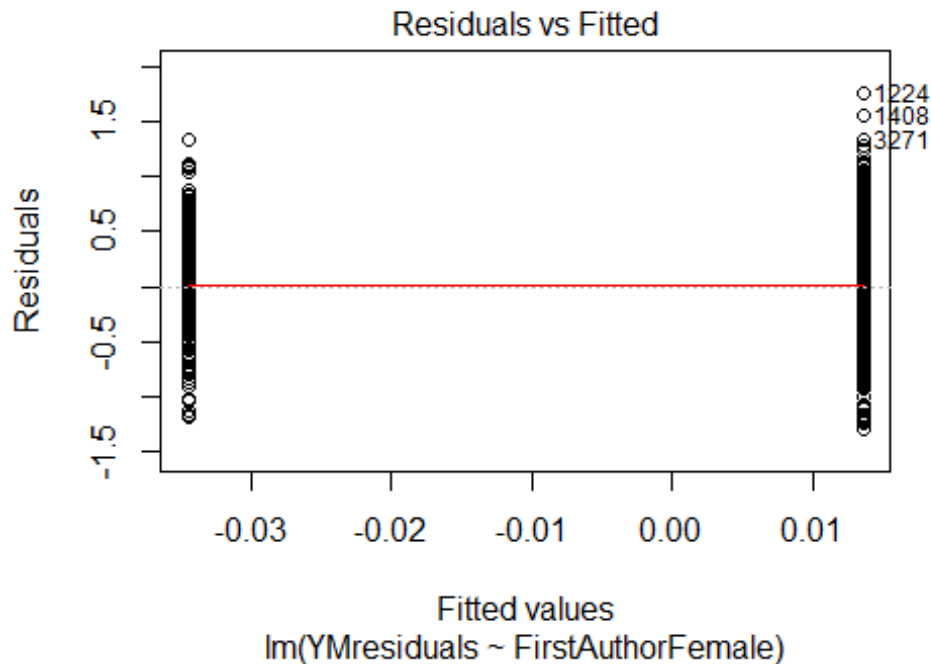
```

```
## 2011 2012
## 106 111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 58 58 43 51 43 58 32 40 56 72 68 92 81 84
## 2011 2012
## 92 97
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 44, df = 16, p-value = 2e-04
```



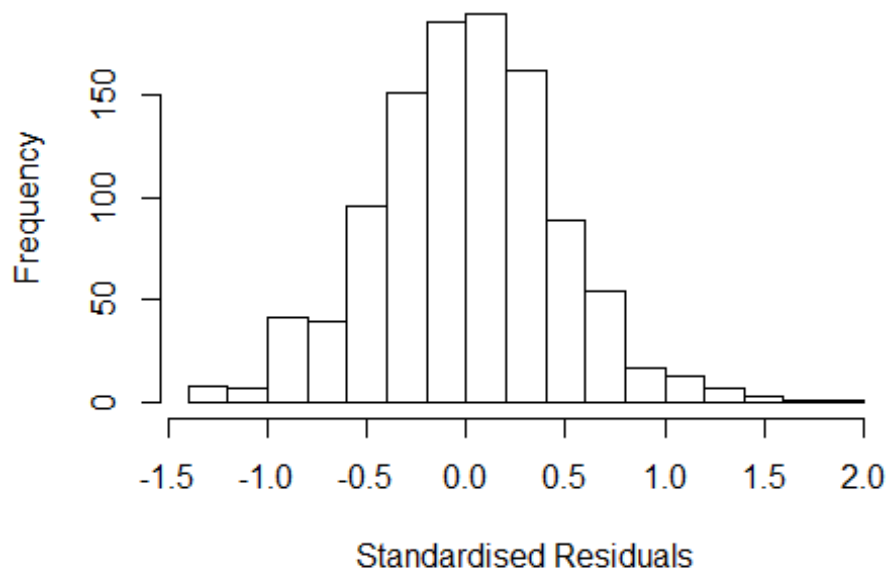
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0085, df = 1, p-value = 0.9
```





```
## [1] "Female first author team size 2018 geometric mean: 3.99893505959367"
## [1] "Male first author team size 2018 geometric mean: 3.56856177059809"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.64916604175671"
## [1] "Male last author team size 2018 geometric mean: 3.74491195346555"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.094 1      1.046
## LastAuthorFemale  1.095 1      1.046
## UniqueAuthors    1.465 4      1.049
## Year             1.682 16      1.016
```

## Residuals from first and last author and team size



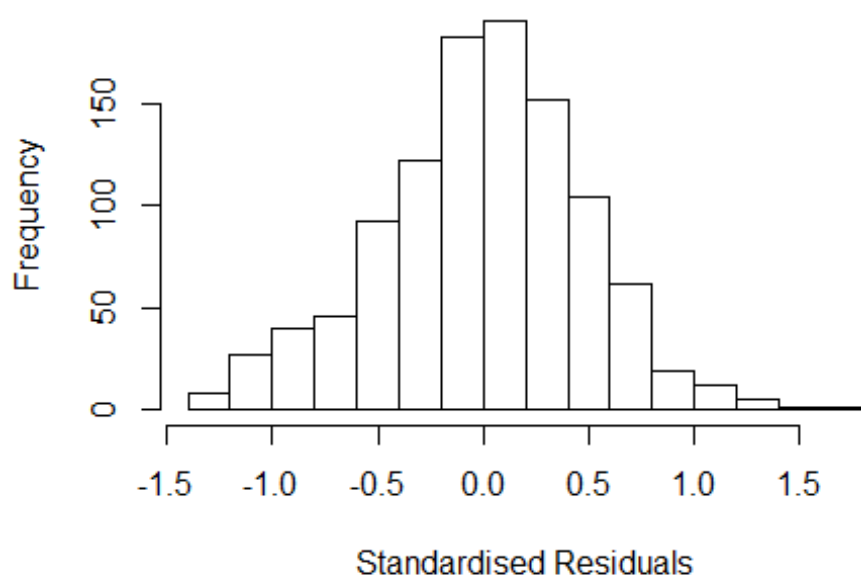
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.36590 -0.27716 0.00419 0.30007 1.81619
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.926540 0.095424 9.71 < 2e-16 ***
## FirstAuthorFemale1 -0.074070 0.032876 -2.25 0.0245 *
## LastAuthorFemale1 0.018926 0.036757 0.51 0.6067
## UniqueAuthors2 0.320579 0.052182 6.14 1.1e-09 ***
## UniqueAuthors3 0.358320 0.052911 6.77 2.1e-11 ***
## UniqueAuthors4 0.379487 0.054659 6.94 6.8e-12 ***
## UniqueAuthors5 0.484406 0.051705 9.37 < 2e-16 ***
## Year1997 -0.000676 0.109205 -0.01 0.9951
## Year1998 0.081044 0.103064 0.79 0.4318
## Year1999 0.053021 0.117448 0.45 0.6518
```

```

## Year2000      -0.004105    0.103557   -0.04    0.9684
## Year2001      -0.009499    0.108289   -0.09    0.9301
## Year2002       0.017272    0.112056    0.15    0.8775
## Year2003      -0.070829    0.111387   -0.64    0.5250
## Year2004      -0.140145    0.108642   -1.29    0.1973
## Year2005      -0.048425    0.099443   -0.49    0.6264
## Year2006      -0.128791    0.105334   -1.22    0.2217
## Year2007      -0.373477    0.116433   -3.21    0.0014 **
## Year2008      -0.116918    0.102395   -1.14    0.2538
## Year2009      -0.144139    0.100311   -1.44    0.1510
## Year2010      -0.065796    0.098884   -0.67    0.5060
## Year2011      -0.201279    0.105230   -1.91    0.0561 .
## Year2012      -0.039367    0.098300   -0.40    0.6889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.439
## Multiple R-squared:  0.144, Adjusted R-squared:  0.126
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 965 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0482 0.8690 0.9510 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.39e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.080 1          1.039
## LastAuthorFemale 1.081 1          1.040
## Year      1.167 16          1.005

```

## Residuals from first and last author



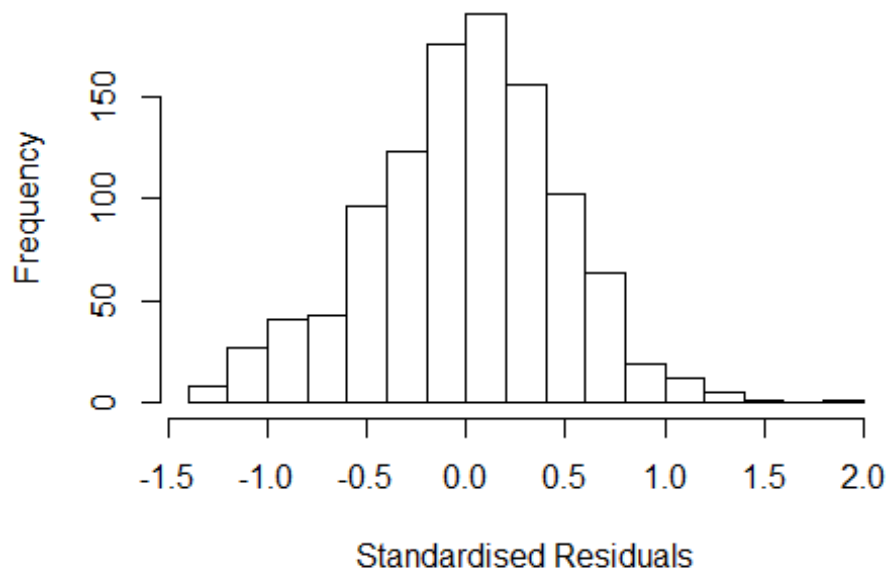
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3032 -0.3049  0.0106  0.3066  1.7839
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16893    0.08986   13.01  <2e-16 ***
## FirstAuthorFemale1 -0.03885    0.03408   -1.14    0.255
## LastAuthorFemale1  0.02951    0.03836    0.77    0.442
## Year1997         0.05147    0.11517    0.45    0.655
## Year1998         0.13428    0.10415    1.29    0.198
## Year1999         0.03825    0.11898    0.32    0.748
## Year2000         0.06610    0.10780    0.61    0.540
## Year2001        -0.00230    0.10799   -0.02    0.983
## Year2002         0.01704    0.12093    0.14    0.888
## Year2003         0.00782    0.10774    0.07    0.942
## Year2004        -0.05104    0.11141   -0.46    0.647
## Year2005         0.03095    0.10117    0.31    0.760
```

```

## Year2006          -0.02367      0.10405      -0.23      0.820
## Year2007          -0.33661      0.12665      -2.66      0.008 **
## Year2008          -0.08132      0.10899      -0.75      0.456
## Year2009          -0.06075      0.10317      -0.59      0.556
## Year2010           0.03424      0.10077       0.34      0.734
## Year2011          -0.11897      0.10818      -1.10      0.272
## Year2012           0.05452      0.10010       0.54      0.586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.047, Adjusted R-squared:  0.0306
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 963 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0915 0.8590 0.9490 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.39e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.08 1      1.039
## Year              1.08 16      1.002

```

## Residuals from first author



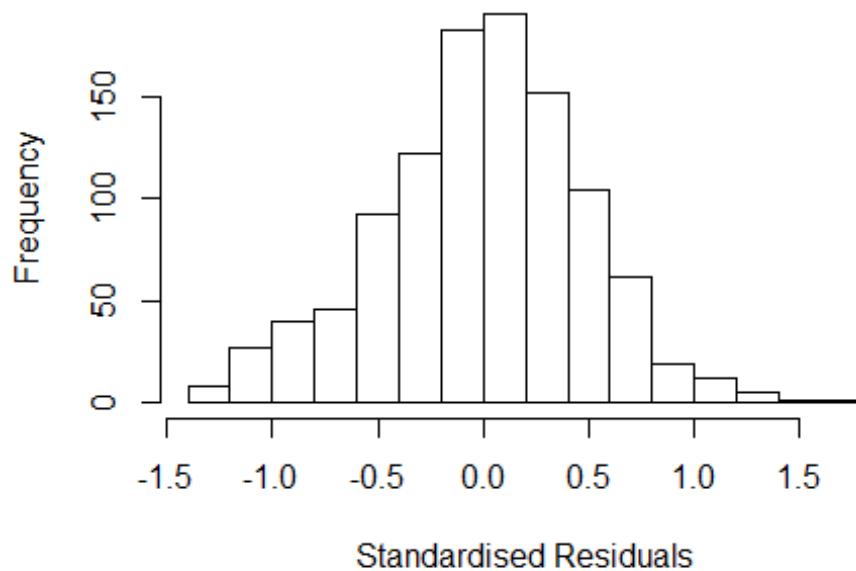
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.307 -0.306 0.013 0.316 1.808
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17400 0.08956 13.11 <2e-16 ***
## FirstAuthorFemale1 -0.03563 0.03427 -1.04 0.2987
## Year1997 0.04980 0.11511 0.43 0.6653
## Year1998 0.13259 0.10427 1.27 0.2038
## Year1999 0.03643 0.11872 0.31 0.7590
## Year2000 0.06532 0.10804 0.60 0.5456
## Year2001 -0.00245 0.10828 -0.02 0.9820
## Year2002 0.01916 0.12094 0.16 0.8742
## Year2003 0.00190 0.10750 0.02 0.9859
## Year2004 -0.05402 0.11146 -0.48 0.6280
## Year2005 0.03079 0.10141 0.30 0.7615
## Year2006 -0.02313 0.10411 -0.22 0.8243
```

```

## Year2007          -0.33451    0.12658   -2.64    0.0083 **
## Year2008          -0.08215    0.10908   -0.75    0.4516
## Year2009          -0.05909    0.10306   -0.57    0.5665
## Year2010           0.03398    0.10086    0.34    0.7363
## Year2011          -0.12205    0.10810   -1.13    0.2591
## Year2012           0.05475    0.10017    0.55    0.5848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.0465, Adjusted R-squared:  0.031
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 93 weights are ~= 1. The remaining 972 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.081  0.859  0.951  0.897  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.39e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.076 1      1.038
## Year              1.076 16      1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2957 -0.2988  0.0132  0.3126  1.7989
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16220    0.09106   12.76  <2e-16 ***
## LastAuthorFemale1 0.02356    0.03874    0.61  0.5432
## Year1997        0.04603    0.11568    0.40  0.6908
## Year1998        0.13347    0.10521    1.27  0.2048
## Year1999        0.04000    0.11987    0.33  0.7387
## Year2000        0.06022    0.10862    0.55  0.5794
## Year2001       -0.00465    0.10835   -0.04  0.9657
## Year2002        0.01819    0.12215    0.15  0.8816
## Year2003        0.00456    0.10898    0.04  0.9666
## Year2004       -0.05857    0.11266   -0.52  0.6033
## Year2005        0.02604    0.10222    0.25  0.7989
## Year2006       -0.02698    0.10502   -0.26  0.7973
```

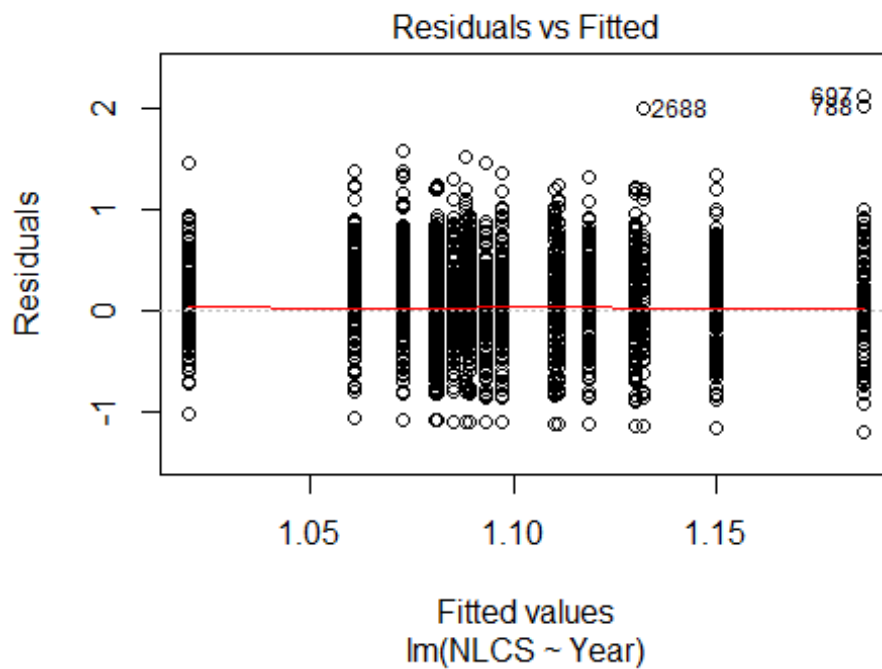


```

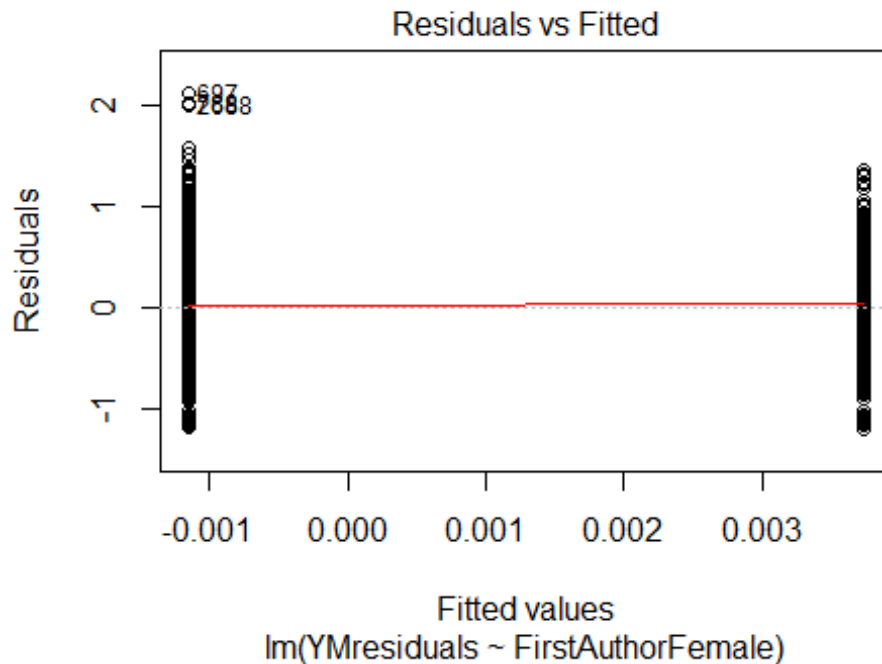
## Year2007          -0.34298      0.12710      -2.70      0.0071 **
## Year2008          -0.08765      0.10960      -0.80      0.4241
## Year2009          -0.06513      0.10399      -0.63      0.5313
## Year2010           0.03104      0.10161       0.31      0.7601
## Year2011          -0.12239      0.10901      -1.12      0.2618
## Year2012           0.05345      0.10118       0.53      0.5974
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.0457, Adjusted R-squared:  0.0302
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 966 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0903 0.8620 0.9510 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.39e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1065"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2312"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 366 394 395 471 440 408 415 370 358 287 257 349 325 350 352
## 2011 2012
## 343 379
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 136 147 155 137 93 171 178 164 156 133 173 181 183 165
## 2011 2012

```

```
## 159 196
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 106 126 138 147 125 86 150 164 155 144 111 152 167 167 149
## 2011 2012
## 143 170
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

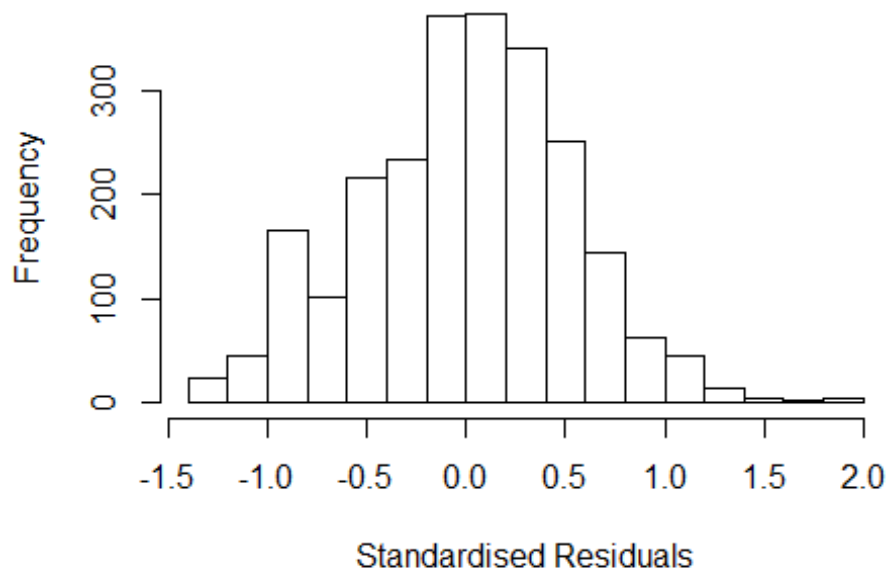


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
```



```
## [1] "Female first author team size 2018 geometric mean: 3.13139144262956"
## [1] "Male first author team size 2018 geometric mean: 3.25510564786095"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.15594026926975"
## [1] "Male last author team size 2018 geometric mean: 3.23539124219294"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.027 1      1.013
## LastAuthorFemale  1.057 1      1.028
## UniqueAuthors     1.171 4      1.020
## Year              1.206 16      1.006
```

## Residuals from first and last author and team size



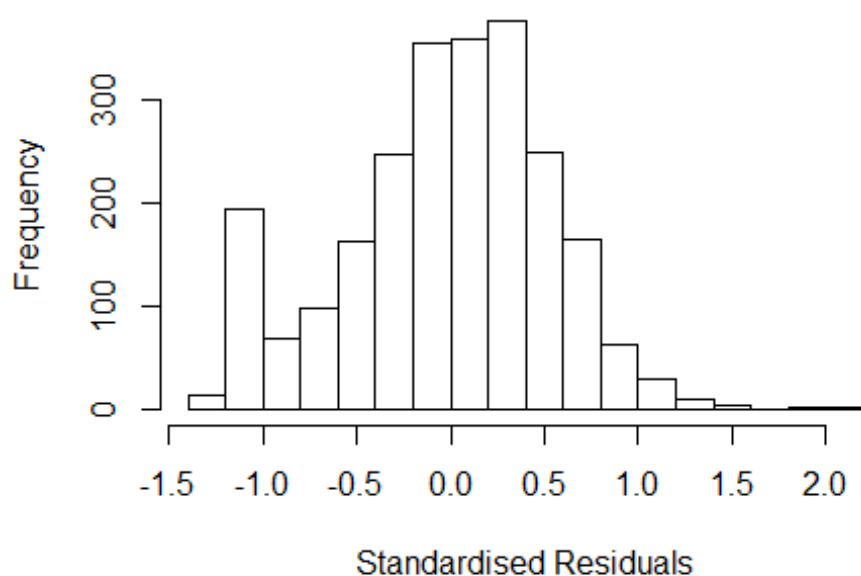
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3485 -0.3444  0.0228  0.3461  1.9972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.90358    0.06959   12.98  <2e-16 ***
## FirstAuthorFemale1 -0.02333    0.02519   -0.93   0.354
## LastAuthorFemale1 -0.05267    0.03042   -1.73   0.083 .
## UniqueAuthors2     0.31382    0.03274    9.59  <2e-16 ***
## UniqueAuthors3     0.31517    0.03587    8.79  <2e-16 ***
## UniqueAuthors4     0.40057    0.03979   10.07  <2e-16 ***
## UniqueAuthors5     0.42481    0.03884   10.94  <2e-16 ***
## Year1997          0.08143    0.08528    0.95   0.340
## Year1998         -0.01736    0.08354   -0.21   0.835
## Year1999         -0.02945    0.08088   -0.36   0.716
```

```

## Year2000      0.02009      0.08199      0.25      0.806
## Year2001     -0.00613      0.09318     -0.07      0.948
## Year2002     -0.04404      0.08376     -0.53      0.599
## Year2003     -0.04367      0.07997     -0.55      0.585
## Year2004     -0.01275      0.07789     -0.16      0.870
## Year2005     -0.05230      0.08027     -0.65      0.515
## Year2006     -0.02746      0.07792     -0.35      0.725
## Year2007     -0.05391      0.07854     -0.69      0.493
## Year2008     -0.02158      0.07589     -0.28      0.776
## Year2009     -0.03634      0.07809     -0.47      0.642
## Year2010     -0.12865      0.07734     -1.66      0.096 .
## Year2011     -0.08008      0.08153     -0.98      0.326
## Year2012     -0.02789      0.07774     -0.36      0.720
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.0903, Adjusted R-squared:  0.0819
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 202 weights are ~= 1. The remaining 2198 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0853 0.8570 0.9470 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## LastAuthorFemale 1.047 1      1.023
## Year      1.071 16      1.002

```

## Residuals from first and last author



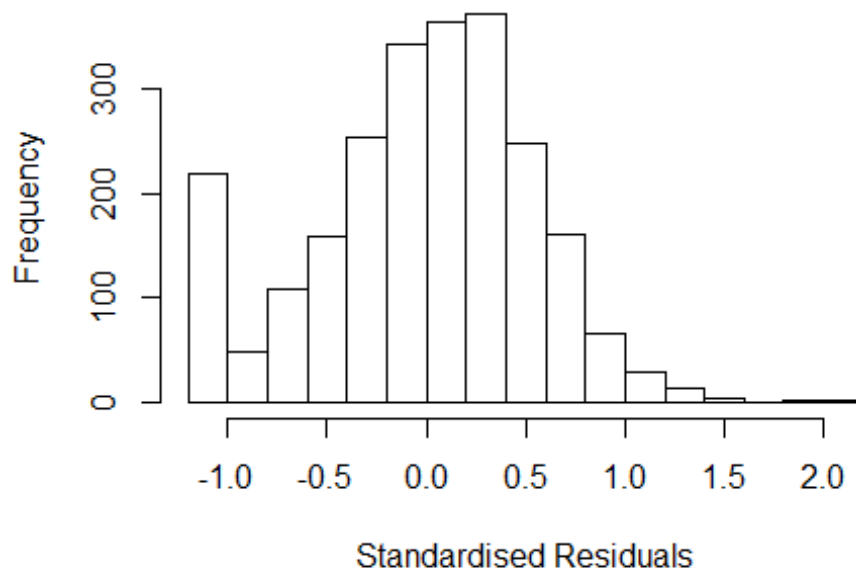
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2168 -0.3497  0.0292  0.3574  2.0902
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11228    0.07110   15.64  <2e-16 ***
## FirstAuthorFemale1  0.01097    0.02588    0.42   0.672
## LastAuthorFemale1 -0.07581    0.03159   -2.40   0.016 *
## Year1997         0.09355    0.09118    1.03   0.305
## Year1998         0.00335    0.08723    0.04   0.969
## Year1999        -0.03089    0.08471   -0.36   0.715
## Year2000         0.05008    0.08742    0.57   0.567
## Year2001         0.02815    0.10206    0.28   0.783
## Year2002        -0.02180    0.08676   -0.25   0.802
## Year2003         0.00835    0.08463    0.10   0.921
## Year2004         0.03756    0.08252    0.46   0.649
## Year2005        -0.00147    0.08565   -0.02   0.986
```

```

## Year2006          0.03496    0.08186    0.43    0.669
## Year2007          -0.01347    0.08245   -0.16    0.870
## Year2008          0.01878    0.08145    0.23    0.818
## Year2009          -0.00386    0.08157   -0.05    0.962
## Year2010          -0.05741    0.08228   -0.70    0.485
## Year2011          -0.02240    0.08805   -0.25    0.799
## Year2012          0.04543    0.08094    0.56    0.575
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.515
## Multiple R-squared:  0.00675,    Adjusted R-squared:  -0.000763
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 215 weights are ~= 1. The remaining 2185 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0616 0.8480 0.9490 0.8940 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1      1.014
## Year              1.029 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1976 -0.3448  0.0281  0.3556  2.0984
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10464    0.07119   15.52  <2e-16 ***
## FirstAuthorFemale1 -0.00235    0.02660   -0.09    0.93
## Year1997         0.09297    0.09143    1.02    0.31
## Year1998         0.00976    0.08707    0.11    0.91
## Year1999        -0.02944    0.08481   -0.35    0.73
## Year2000         0.04831    0.08771    0.55    0.58
## Year2001         0.02485    0.10236    0.24    0.81
## Year2002        -0.02624    0.08699   -0.30    0.76
## Year2003         0.00708    0.08480    0.08    0.93
## Year2004         0.03947    0.08269    0.48    0.63
## Year2005         0.00302    0.08570    0.04    0.97
## Year2006         0.03242    0.08179    0.40    0.69
```

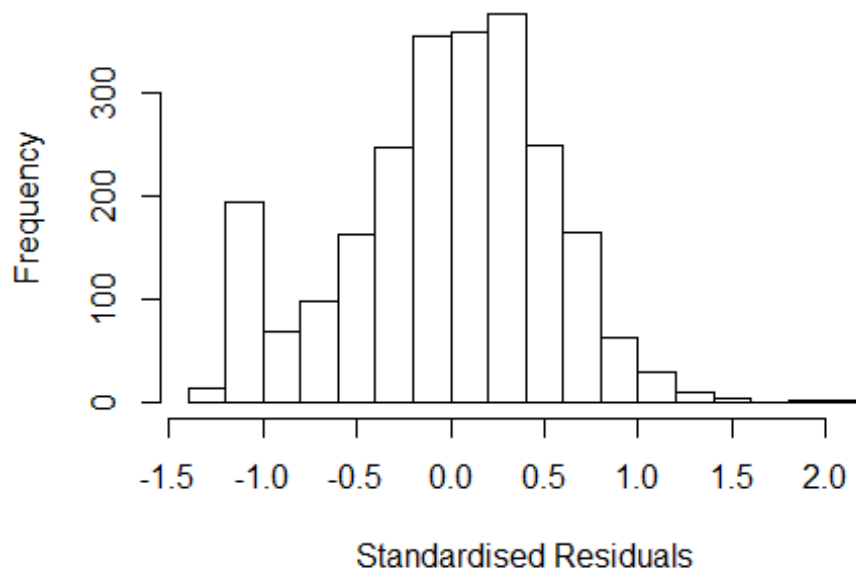


```

## Year2007      -0.01737    0.08262   -0.21    0.83
## Year2008      0.01504    0.08174    0.18    0.85
## Year2009     -0.00566    0.08175   -0.07    0.94
## Year2010     -0.06344    0.08284   -0.77    0.44
## Year2011     -0.02821    0.08826   -0.32    0.75
## Year2012      0.04165    0.08114    0.51    0.61
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.00423,    Adjusted R-squared:  -0.00288
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 218 weights are ~= 1. The remaining 2182 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0583 0.8520 0.9480 0.8940 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.047 1          1.023
## Year            1.047 16          1.001

```

## Residuals from last author



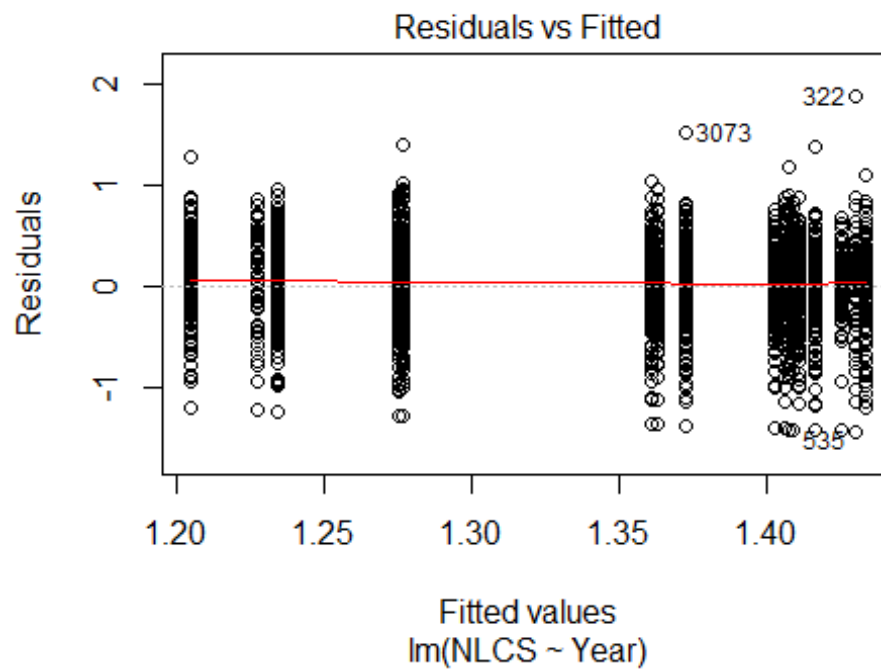
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2073 -0.3479 0.0281 0.3539 2.0887
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11380 0.07083 15.73 <2e-16 ***
## LastAuthorFemale1 -0.07318 0.03211 -2.28 0.023 *
## Year1997 0.09355 0.09118 1.03 0.305
## Year1998 0.00458 0.08722 0.05 0.958
## Year1999 -0.03133 0.08464 -0.37 0.711
## Year2000 0.05074 0.08744 0.58 0.562
## Year2001 0.02803 0.10204 0.27 0.784
## Year2002 -0.02100 0.08672 -0.24 0.809
## Year2003 0.00913 0.08460 0.11 0.914
## Year2004 0.03779 0.08252 0.46 0.647
## Year2005 -0.00109 0.08569 -0.01 0.990
## Year2006 0.03554 0.08190 0.43 0.664
```

```

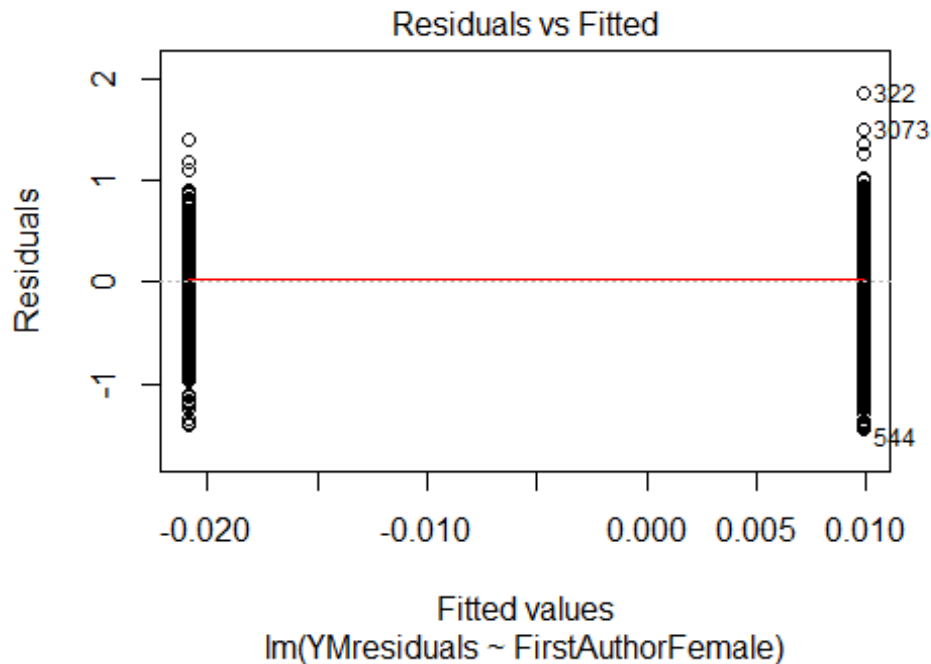
## Year2007          -0.01233      0.08245    -0.15      0.881
## Year2008           0.02013      0.08148      0.25      0.805
## Year2009          -0.00260      0.08166     -0.03      0.975
## Year2010          -0.05628      0.08235     -0.68      0.494
## Year2011          -0.02140      0.08799     -0.24      0.808
## Year2012           0.04715      0.08090      0.58      0.560
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.00672,    Adjusted R-squared:  -0.000367
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 213 weights are ~= 1. The remaining 2187 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.062  0.848  0.949  0.895  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2400"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2400"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 232 240 243 245 232 243 260 213 218 163 228 217 168 230 210
## 2011 2012
## 246 277
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 66 69 134 104 62 173 146 144 120 170 154 117 164 142
## 2011 2012

```

```
## 177 209
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 59 62 63 115 96 51 153 133 130 103 150 138 100 149 130
## 2011 2012
## 165 188
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 89, df = 16, p-value = 4e-12
```

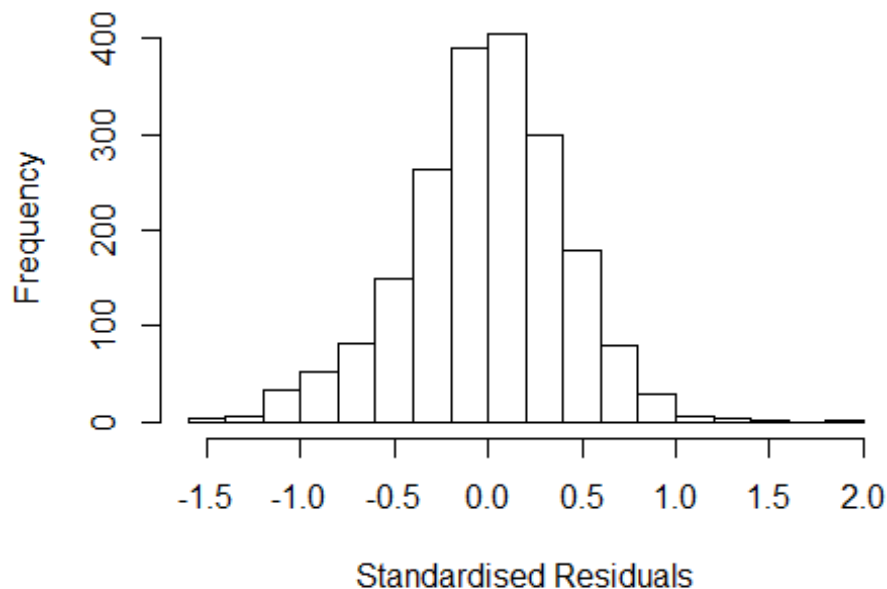


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.048, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 3.73989832109917"
## [1] "Male first author team size 2018 geometric mean: 3.35143091702537"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6500, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.17412806315592"
## [1] "Male last author team size 2018 geometric mean: 3.64240246195524"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.119 1 1.058
## LastAuthorFemale 1.033 1 1.017
## UniqueAuthors 1.284 4 1.032
## Year 1.319 16 1.009
```

## Residuals from first and last author and team size



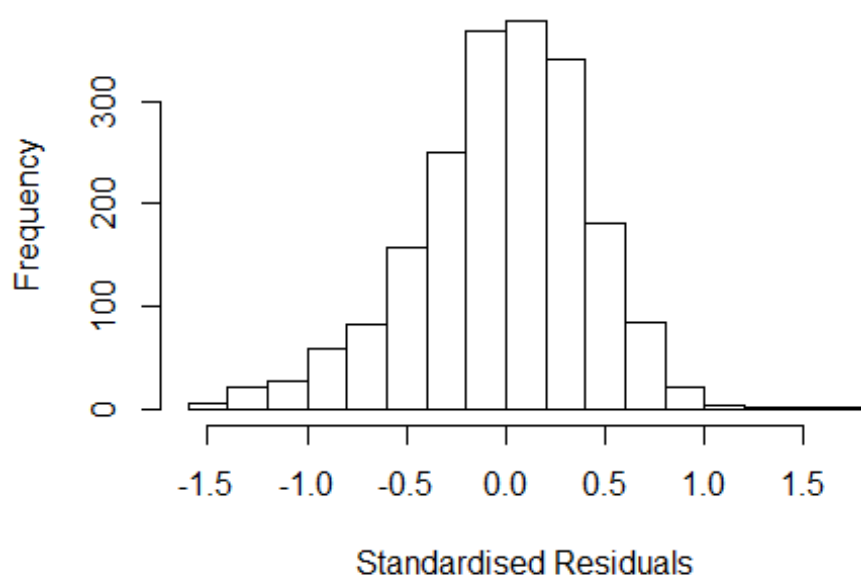
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.52386 -0.26181 0.00526 0.25829 1.80460
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23131 0.08140 15.13 < 2e-16 ***
## FirstAuthorFemale1 -0.05860 0.02110 -2.78 0.0055 **
## LastAuthorFemale1 -0.03756 0.02505 -1.50 0.1339
## UniqueAuthors2 0.20490 0.03950 5.19 2.3e-07 ***
## UniqueAuthors3 0.23436 0.04007 5.85 5.8e-09 ***
## UniqueAuthors4 0.28543 0.04215 6.77 1.7e-11 ***
## UniqueAuthors5 0.33097 0.03724 8.89 < 2e-16 ***
## Year1997 0.05819 0.09378 0.62 0.5350
## Year1998 -0.00472 0.09821 -0.05 0.9616
## Year1999 -0.01025 0.08690 -0.12 0.9061
```

```

## Year2000      0.00571    0.08309    0.07    0.9452
## Year2001     -0.20169    0.10973   -1.84    0.0662 .
## Year2002     -0.00458    0.08282   -0.06    0.9559
## Year2003     -0.05177    0.08715   -0.59    0.5526
## Year2004     -0.01523    0.08379   -0.18    0.8557
## Year2005     -0.02147    0.08529   -0.25    0.8013
## Year2006     -0.05624    0.08492   -0.66    0.5079
## Year2007     -0.06288    0.08394   -0.75    0.4539
## Year2008      0.03838    0.08791    0.44    0.6625
## Year2009     -0.12240    0.08658   -1.41    0.1576
## Year2010     -0.19265    0.08951   -2.15    0.0315 *
## Year2011     -0.16078    0.08594   -1.87    0.0615 .
## Year2012     -0.16560    0.08574   -1.93    0.0536 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.393
## Multiple R-squared:  0.0944, Adjusted R-squared:  0.0842
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 1807 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0015 0.8650 0.9500 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.04e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.075 1 1.037
## LastAuthorFemale 1.025 1 1.013
## Year 1.081 16 1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4997 -0.2754  0.0144  0.2646  1.7993
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.423102   0.085188   16.71  <2e-16 ***
## FirstAuthorFemale1 -0.019272   0.021107   -0.91    0.36
## LastAuthorFemale1 -0.028534   0.025563   -1.12    0.26
## Year1997         0.076623   0.102891    0.74    0.46
## Year1998         0.010708   0.104882    0.10    0.92
## Year1999         0.017048   0.093626    0.18    0.86
## Year2000         0.031461   0.089615    0.35    0.73
## Year2001        -0.144192   0.115303   -1.25    0.21
## Year2002         0.025521   0.089650    0.28    0.78
## Year2003        -0.031722   0.092585   -0.34    0.73
## Year2004         0.000534   0.090393    0.01    1.00
## Year2005         0.025103   0.091252    0.28    0.78
```

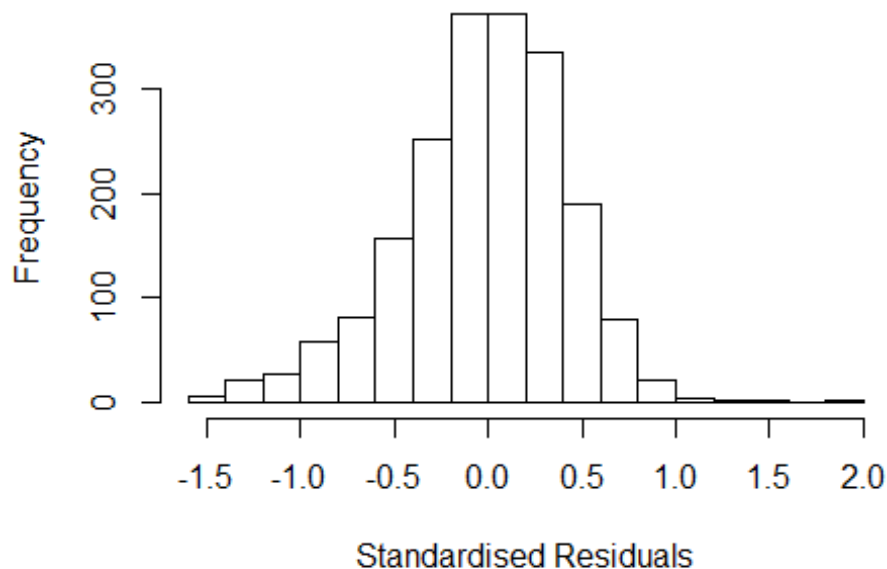


```

## Year2006          -0.016455    0.092024    -0.18      0.86
## Year2007          -0.026676    0.090588    -0.29      0.77
## Year2008           0.058889    0.095970     0.61      0.54
## Year2009          -0.113709    0.094796    -1.20      0.23
## Year2010          -0.158236    0.096860    -1.63      0.10
## Year2011          -0.133720    0.093634    -1.43      0.15
## Year2012          -0.141049    0.093142    -1.51      0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0328, Adjusted R-squared:  0.024
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1828 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0061 0.8630 0.9510 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1      1.032
## Year              1.065 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4930 -0.2738 0.0123 0.2646 1.8060
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.41833 0.08507 16.67 <2e-16 ***
## FirstAuthorFemale1 -0.02263 0.02106 -1.07 0.28
## Year1997 0.07462 0.10281 0.73 0.47
## Year1998 0.01189 0.10496 0.11 0.91
## Year1999 0.01898 0.09366 0.20 0.84
## Year2000 0.03302 0.08969 0.37 0.71
## Year2001 -0.14170 0.11527 -1.23 0.22
## Year2002 0.02817 0.08965 0.31 0.75
## Year2003 -0.03048 0.09268 -0.33 0.74
## Year2004 0.00089 0.09043 0.01 0.99
## Year2005 0.02679 0.09126 0.29 0.77
## Year2006 -0.01637 0.09202 -0.18 0.86
```

```

## Year2007          -0.02725    0.09062   -0.30    0.76
## Year2008           0.06078    0.09608    0.63    0.53
## Year2009          -0.11399    0.09477   -1.20    0.23
## Year2010          -0.15790    0.09709   -1.63    0.10
## Year2011          -0.13452    0.09365   -1.44    0.15
## Year2012          -0.14185    0.09317   -1.52    0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0321, Adjusted R-squared:  0.0237
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~ = 1. The remaining 1823 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0051 0.8640 0.9500 0.8930 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1          1.008
## Year            1.016 16          1.001
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.4964 -0.2737  0.0143  0.2650  1.8026

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.42e+00  8.49e-02  16.72  <2e-16 ***
## LastAuthorFemale1 -3.18e-02  2.55e-02  -1.25   0.212
## Year1997       7.64e-02  1.02e-01   0.75   0.456
## Year1998       9.70e-03  1.05e-01   0.09   0.926
## Year1999       1.54e-02  9.31e-02   0.16   0.869
## Year2000       2.81e-02  8.90e-02   0.32   0.752
## Year2001      -1.46e-01  1.15e-01  -1.27   0.205
## Year2002       2.28e-02  8.91e-02   0.26   0.798
## Year2003      -3.48e-02  9.20e-02  -0.38   0.706
## Year2004      -7.99e-05  9.00e-02   0.00   0.999
## Year2005       2.27e-02  9.07e-02   0.25   0.803
## Year2006      -1.99e-02  9.14e-02  -0.22   0.828
## Year2007      -2.93e-02  9.01e-02  -0.33   0.745
## Year2008       5.56e-02  9.53e-02   0.58   0.560
## Year2009      -1.15e-01  9.44e-02  -1.22   0.223
## Year2010      -1.63e-01  9.62e-02  -1.69   0.091 .
## Year2011      -1.36e-01  9.32e-02  -1.46   0.145
## Year2012      -1.45e-01  9.24e-02  -1.57   0.117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0323, Adjusted R-squared:  0.024
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1828 ones are summarized as
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## 0.0059 0.8630 0.9510 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1985"
## [1] ""

```

```

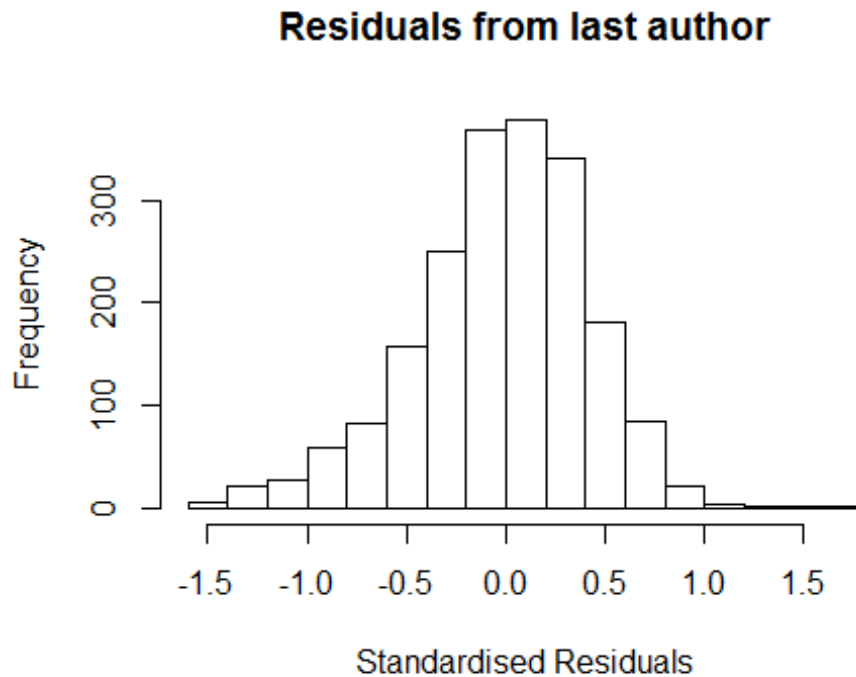
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2401"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2008 2009 2010 2011 2012
##    1    6    6   13    5
##
## 2008 2009 2010 2011 2012
##    1    5    5   12    4
##
## 2008 2009 2010 2011 2012
##    1    5    3    9    3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 6.83617479166468"
## [1] "Male first author team size 2018 geometric mean: 6.16014057648204"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.48074069840786"
## [1] "Male last author team size 2018 geometric mean: 6.49223611101639"

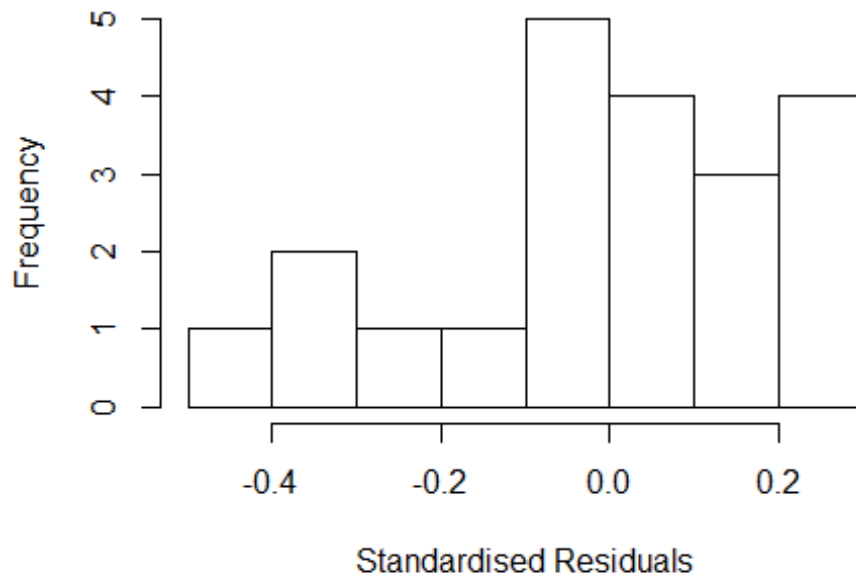
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5.5, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.435e+00 1          1.853
## LastAuthorFemale  6.962e+00 1          2.639
## UniqueAuthors    1.699e+15 3          345.429
## Year              2.988e+15 4          85.986
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.43995 -0.07557 0.00145 0.14958 0.25930
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7675 0.2518 3.05 0.0111 *
## FirstAuthorFemale1 -0.2108 0.1063 -1.98 0.0729 .
## LastAuthorFemale1 0.0961 0.0995 0.97 0.3547
## UniqueAuthors3 -0.1596 0.2056 -0.78 0.4541
## UniqueAuthors4 -0.0380 0.0966 -0.39 0.7017
## UniqueAuthors5 -0.1160 0.1113 -1.04 0.3198
## Year2009 0.6748 0.3123 2.16 0.0537 .
## Year2010 0.6571 0.2846 2.31 0.0414 *
## Year2011 0.6794 0.2056 3.30 0.0070 **
## Year2012 1.0780 0.3012 3.58 0.0043 **
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.294
## Multiple R-squared:  0.513, Adjusted R-squared:  0.114
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 17 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.806  0.930  0.975  0.950  0.994  0.998
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           4.76e-03           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.428 1           1.852
## LastAuthorFemale  6.520 1           2.553
## Year              15.034 4           1.403
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5684 -0.0982  0.0116  0.1298  0.2519
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5607    0.0927     6.05 3.0e-05 ***

```



```

## FirstAuthorFemale1 -0.2181      0.1465    -1.49      0.16
## LastAuthorFemale1  0.1433      0.0927     1.55      0.14
## Year2009            0.8106      0.1160     6.99    6.4e-06 ***
## Year2010            0.7786      0.1391     5.60    6.6e-05 ***
## Year2011            0.7784      0.0641    12.14    8.0e-09 ***
## Year2012            1.1835      0.2034     5.82    4.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.182
## Multiple R-squared:  0.564, Adjusted R-squared:  0.377
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 3 weights are ~= 1. The remaining 18 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.845   0.945   0.878   0.990   0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.76e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

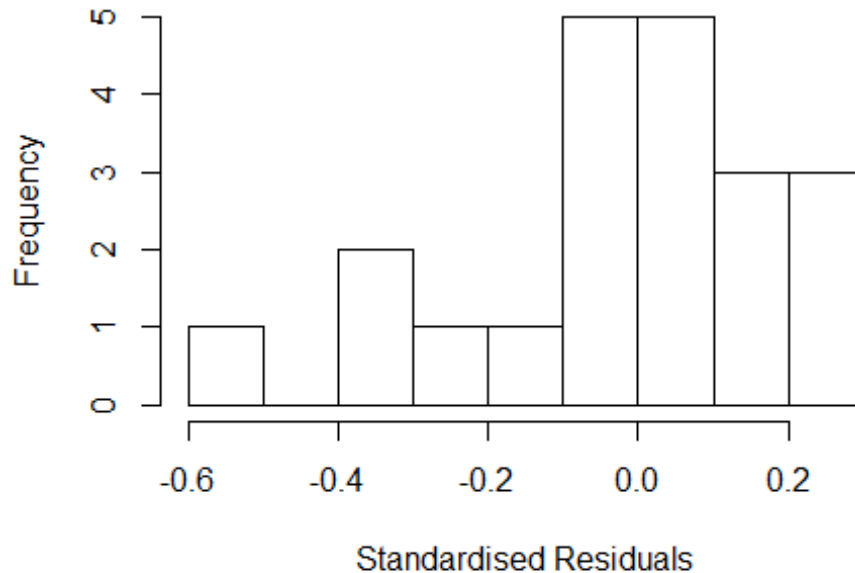
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

```

```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

### Residuals from first and last author



```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 54.24 1          7.365
## Year              54.24 4          1.647

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5355 -0.1394  0.0275  0.0985  0.2070
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    7.04e-01   2.41e-09  2.93e+08  <2e-16 ***
## FirstAuthorFemale1 -1.84e-01   5.84e-01 -3.10e-01  0.7574
```

```

## Year2009          7.43e-01   1.91e-01   3.89e+00   0.0014 **
## Year2010          6.13e-01   3.46e-01   1.77e+00   0.0969 .
## Year2011          7.22e-01   2.60e-01   2.78e+00   0.0141 *
## Year2012          1.05e+00   3.33e-01   3.16e+00   0.0065 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.146
## Multiple R-squared:  0.609, Adjusted R-squared:  0.479
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 2 weights are ~= 1. The remaining 19 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.153  0.811  0.943  0.819  0.973  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.76e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

```

```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

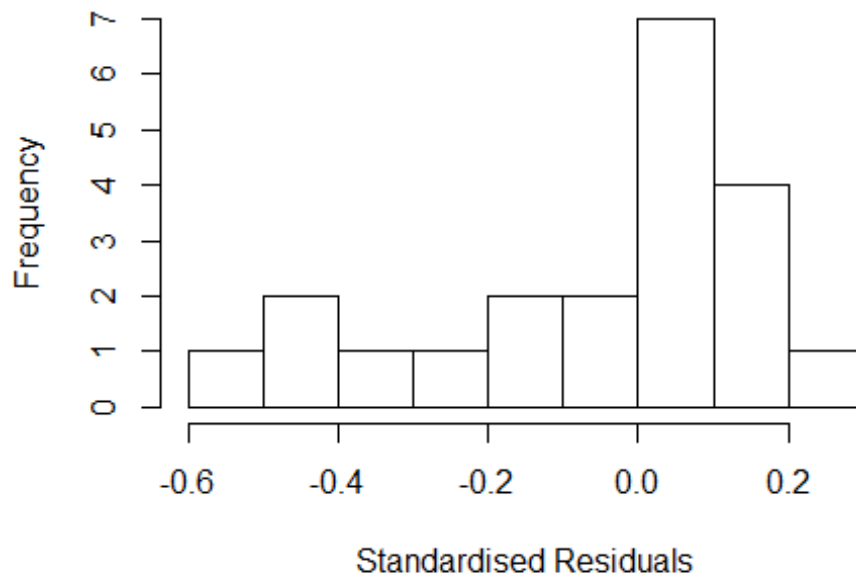
```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

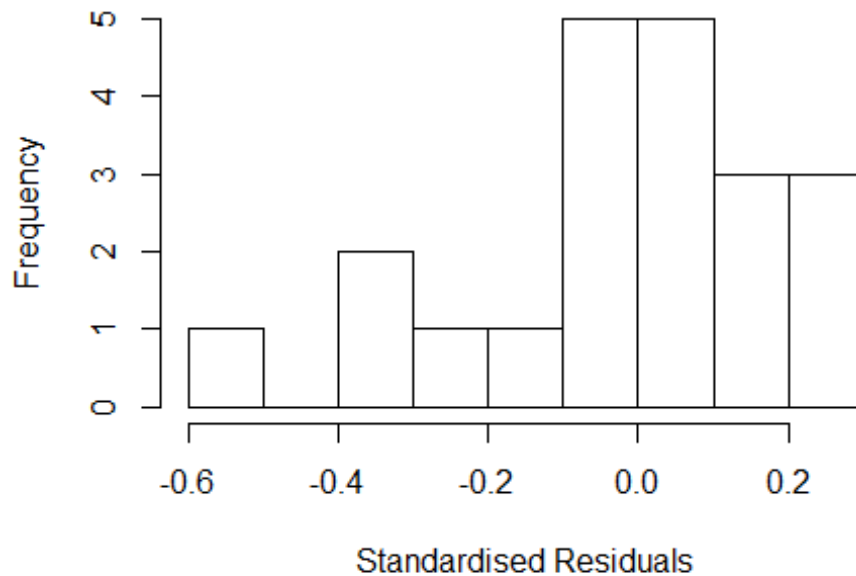
```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

### Residuals from first author



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 40.52 1      6.365
## Year             40.52 4      1.588
```

## Residuals from last author



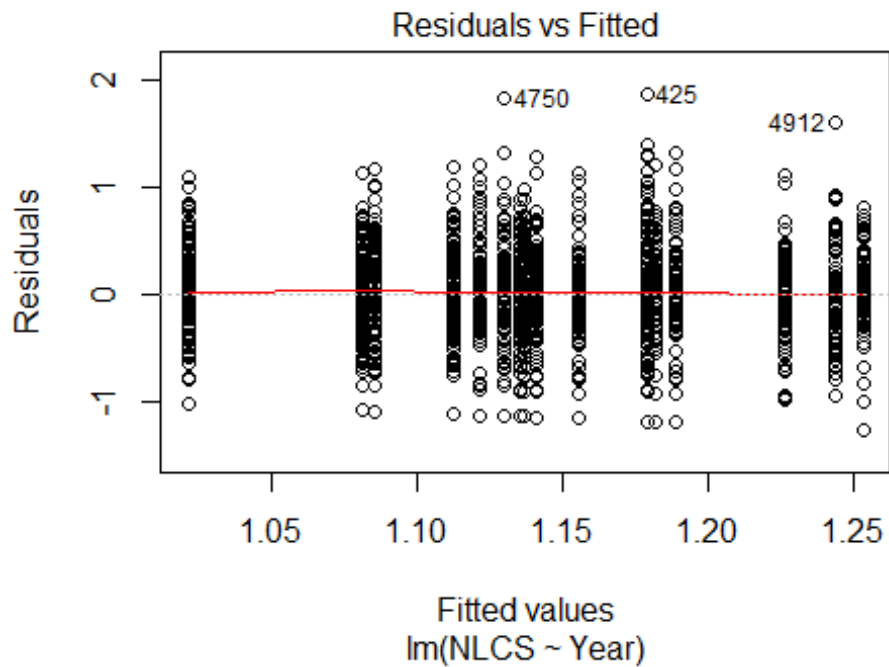
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.6720 -0.1302 0.0213 0.0603 0.2128
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.562 0.323 1.74 0.10255
## LastAuthorFemale1 0.142 0.323 0.44 0.66539
## Year2009 0.696 0.135 5.17 0.00011 ***
## Year2010 0.639 0.284 2.25 0.04016 *
## Year2011 0.769 0.132 5.84 3.2e-05 ***
## Year2012 1.167 0.367 3.18 0.00625 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.147
## Multiple R-squared: 0.683, Adjusted R-squared: 0.578
## Convergence in 21 IRWLS iterations
```

```

##
## Robustness weights:
## observation 5 is an outlier with |weight| <= 0.0019 ( < 0.0048);
## one weight is ~= 1. The remaining 19 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.174  0.887  0.954   0.854   0.989   0.998
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           4.76e-03           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 21"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2402"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   379  303  346  311  362  287  264  221  231  183  203  217  207  128  147
## 2011 2012
##   157  123
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   132   94   98  123  100  102  110  102  116  105  111  106  115   71   81
## 2011 2012
##    96   78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   119   83   92  110   88   92   94   93  100   89   97   89   96   64   69
## 2011 2012
##    83   67
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##

```

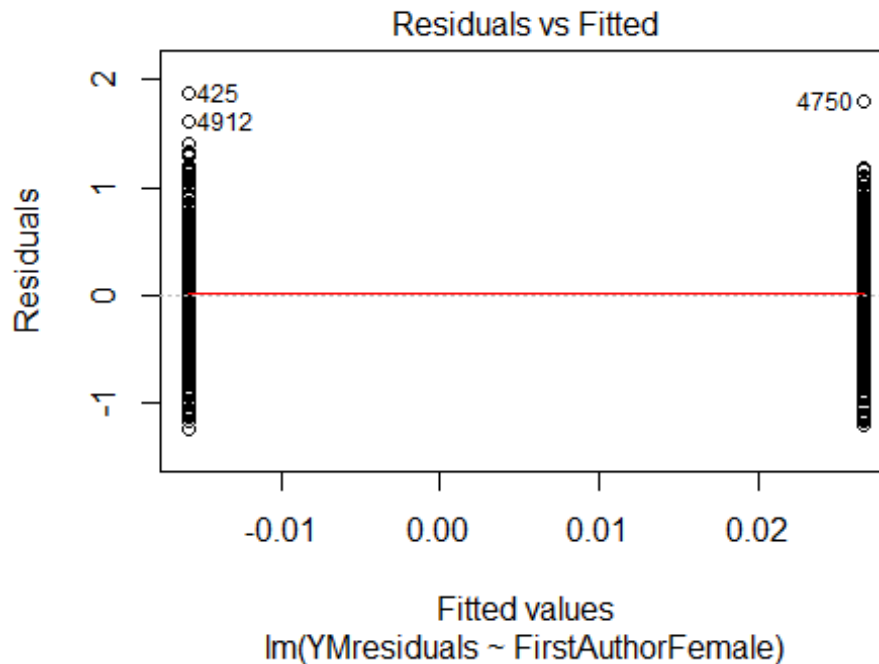
```
## data: NLCS by Year
## Bartlett's K-squared = 75, df = 16, p-value = 1e-09
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.89, df = 1, p-value = 0.3

## [1] "Female first author team size 2018 geometric mean: 3.73025190546966"
## [1] "Male first author team size 2018 geometric mean: 3.86317271219082"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

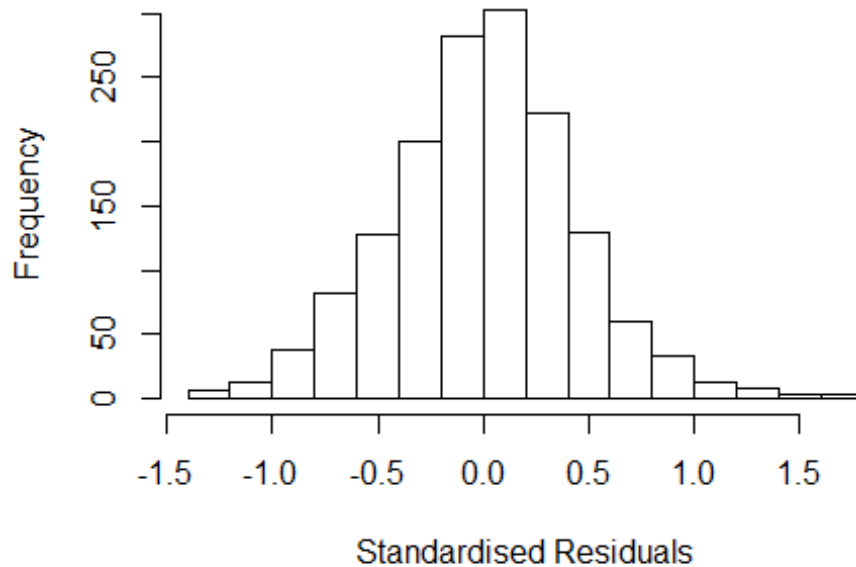


```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 580, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.04136718649179"
## [1] "Male last author team size 2018 geometric mean: 4.09849613333231"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.067	1	1.033
LastAuthorFemale	1.083	1	1.041
UniqueAuthors	1.427	4	1.045
Year	1.492	16	1.013



## Residuals from first and last author and team size



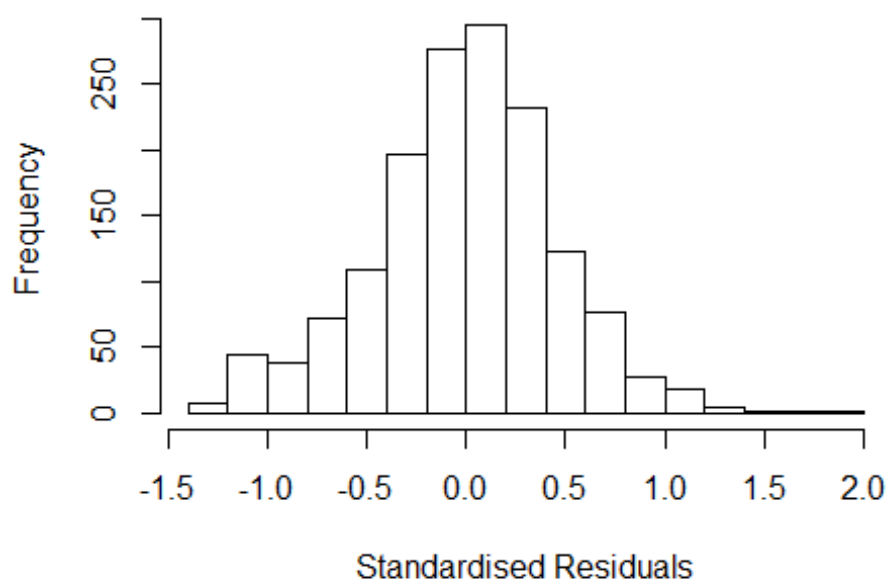
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37796 -0.26083  0.00959  0.27676  1.71136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8547    0.0783   10.92 < 2e-16 ***
## FirstAuthorFemale1  0.0113    0.0242    0.47  0.639
## LastAuthorFemale1  0.0121    0.0307    0.39  0.693
## UniqueAuthors2     0.2587    0.0595    4.35 1.5e-05 ***
## UniqueAuthors3     0.3862    0.0584    6.61 5.3e-11 ***
## UniqueAuthors4     0.3941    0.0576    6.84 1.1e-11 ***
## UniqueAuthors5     0.5513    0.0559    9.85 < 2e-16 ***
## Year1997          -0.0280    0.0812   -0.35  0.730
## Year1998          -0.0337    0.0842   -0.40  0.689
## Year1999          -0.1299    0.0839   -1.55  0.121
```

```

## Year2000          -0.0927      0.0847   -1.09    0.274
## Year2001          -0.1187      0.0834   -1.42    0.155
## Year2002          -0.1825      0.0849   -2.15    0.032 *
## Year2003          -0.0956      0.0832   -1.15    0.251
## Year2004          -0.1192      0.0816   -1.46    0.144
## Year2005          -0.1206      0.0777   -1.55    0.121
## Year2006          -0.1087      0.0769   -1.41    0.158
## Year2007          -0.1196      0.0791   -1.51    0.131
## Year2008          -0.0339      0.0761   -0.45    0.656
## Year2009          -0.0518      0.0829   -0.63    0.532
## Year2010          -0.0709      0.0833   -0.85    0.394
## Year2011          -0.1777      0.0845   -2.10    0.036 *
## Year2012          -0.1012      0.0881   -1.15    0.251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.135, Adjusted R-squared:  0.122
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1406 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0333 0.8600 0.9510 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.56e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1          1.016
## LastAuthorFemale 1.055 1          1.027
## Year 1.089 16          1.003

```

## Residuals from first and last author



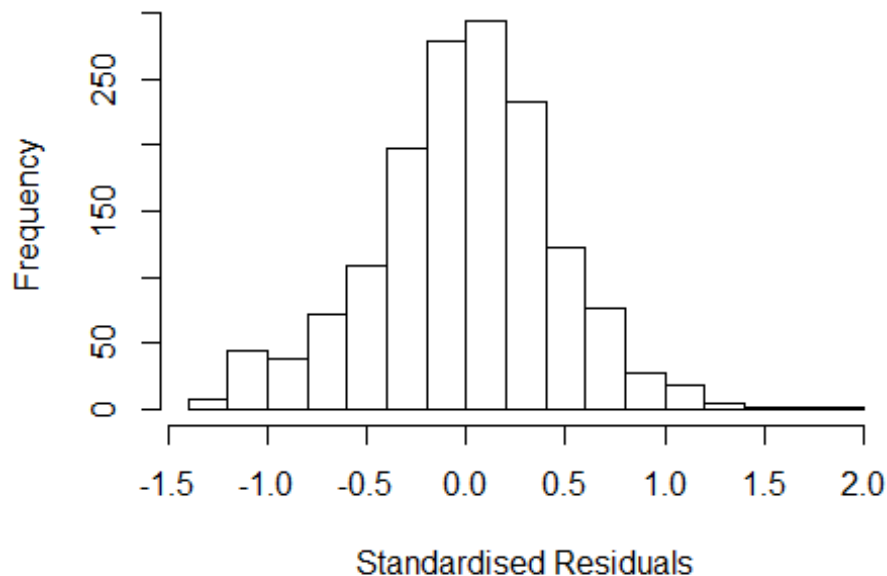
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2557 -0.2805  0.0105  0.2728  1.8840
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.152011   0.074856  15.39  <2e-16 ***
## FirstAuthorFemale1  0.053371   0.024676   2.16   0.031 *
## LastAuthorFemale1  0.001837   0.031700   0.06   0.954
## Year1997        -0.018327   0.090585  -0.20   0.840
## Year1998         0.020035   0.088905   0.23   0.822
## Year1999        -0.085520   0.090518  -0.94   0.345
## Year2000        -0.072562   0.089391  -0.81   0.417
## Year2001        -0.037009   0.088118  -0.42   0.675
## Year2002        -0.131805   0.092041  -1.43   0.152
## Year2003        -0.008698   0.089563  -0.10   0.923
## Year2004        -0.054742   0.087072  -0.63   0.530
## Year2005        -0.048540   0.084230  -0.58   0.565
```

```

## Year2006          0.000536    0.082738    0.01    0.995
## Year2007         -0.050361    0.085917   -0.59    0.558
## Year2008          0.061435    0.081586    0.75    0.452
## Year2009          0.052376    0.087723    0.60    0.551
## Year2010          0.048488    0.090233    0.54    0.591
## Year2011         -0.047584    0.091177   -0.52    0.602
## Year2012          0.009311    0.094961    0.10    0.922
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.0041
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1386 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0086 0.8560 0.9530 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1      1.016
## Year              1.033 16      1.001

```

## Residuals from first author



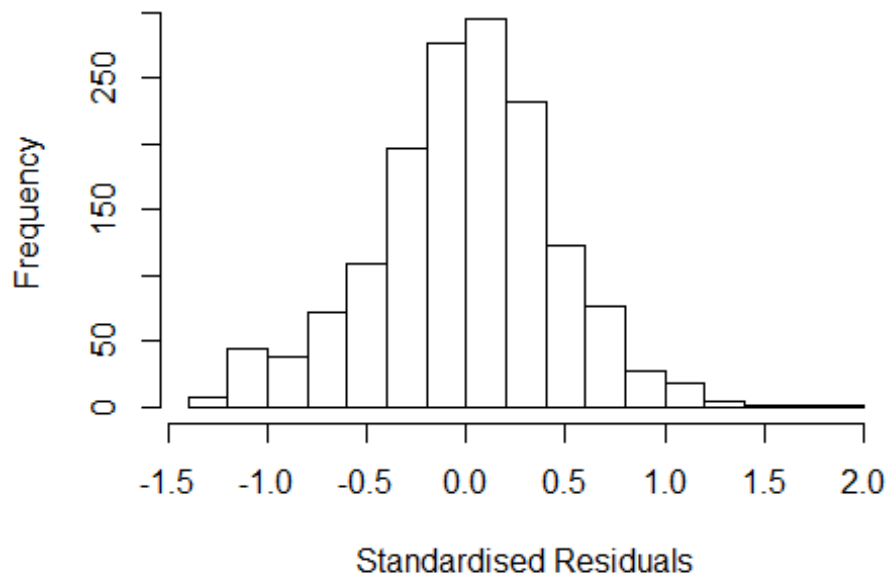
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2545 -0.2807 0.0103 0.2728 1.8838
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.152237 0.074585 15.45 <2e-16 ***
## FirstAuthorFemale1 0.053496 0.024794 2.16 0.031 *
## Year1997 -0.018478 0.090534 -0.20 0.838
## Year1998 0.019957 0.088841 0.22 0.822
## Year1999 -0.085585 0.090470 -0.95 0.344
## Year2000 -0.072551 0.089388 -0.81 0.417
## Year2001 -0.037027 0.088097 -0.42 0.674
## Year2002 -0.131647 0.091987 -1.43 0.153
## Year2003 -0.008711 0.089545 -0.10 0.923
## Year2004 -0.054806 0.086985 -0.63 0.529
## Year2005 -0.048535 0.084210 -0.58 0.564
## Year2006 0.000466 0.082692 0.01 0.996
```

```

## Year2007      -0.050268    0.085876   -0.59    0.558
## Year2008      0.061562    0.081585    0.75    0.451
## Year2009      0.052479    0.087669    0.60    0.550
## Year2010      0.048722    0.090194    0.54    0.589
## Year2011     -0.047502    0.091237   -0.52    0.603
## Year2012      0.009459    0.094891    0.10    0.921
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.00475
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 139 weights are ~= 1. The remaining 1386 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0087 0.8570 0.9530 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.055 1          1.027
## Year            1.055 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2269 -0.2754 0.0109 0.2738 1.8686
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16739 0.07467 15.63 <2e-16 ***
## LastAuthorFemale1 0.00813 0.03184 0.26 0.80
## Year1997 -0.01608 0.09052 -0.18 0.86
## Year1998 0.02336 0.08854 0.26 0.79
## Year1999 -0.08324 0.09025 -0.92 0.36
## Year2000 -0.07380 0.08935 -0.83 0.41
## Year2001 -0.03289 0.08824 -0.37 0.71
## Year2002 -0.12816 0.09208 -1.39 0.16
## Year2003 -0.00273 0.08913 -0.03 0.98
## Year2004 -0.04602 0.08673 -0.53 0.60
## Year2005 -0.04240 0.08397 -0.50 0.61
## Year2006 0.00566 0.08244 0.07 0.95
```

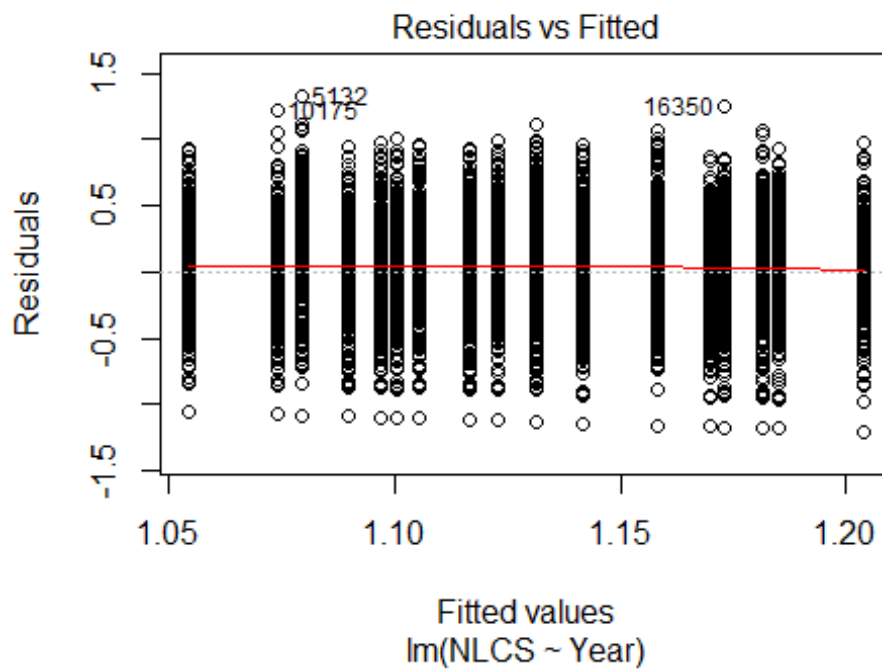
```

## Year2007          -0.04972      0.08579      -0.58      0.56
## Year2008           0.06271      0.08142       0.77      0.44
## Year2009           0.05949      0.08747       0.68      0.50
## Year2010           0.04786      0.09070       0.53      0.60
## Year2011          -0.04542      0.09125      -0.50      0.62
## Year2012           0.01151      0.09553       0.12      0.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.0127, Adjusted R-squared:  0.00152
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1391 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0118 0.8570 0.9520 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.56e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1525"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2403"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 907 838 931 893 977 1071 1005 734 764 793 772 779 741 598 620
## 2011 2012
## 607 617
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 382 339 454 439 368 266 598 421 421 498 461 504 471 382 419
## 2011 2012

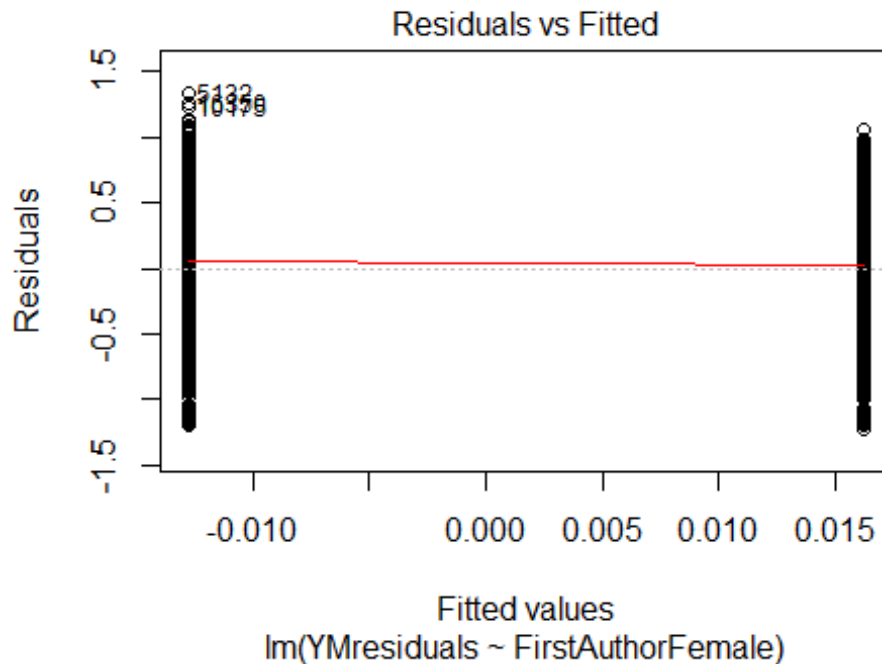
```



```
## 416 444
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 339 295 403 373 329 234 519 349 369 409 400 435 404 335 353
## 2011 2012
## 365 386
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

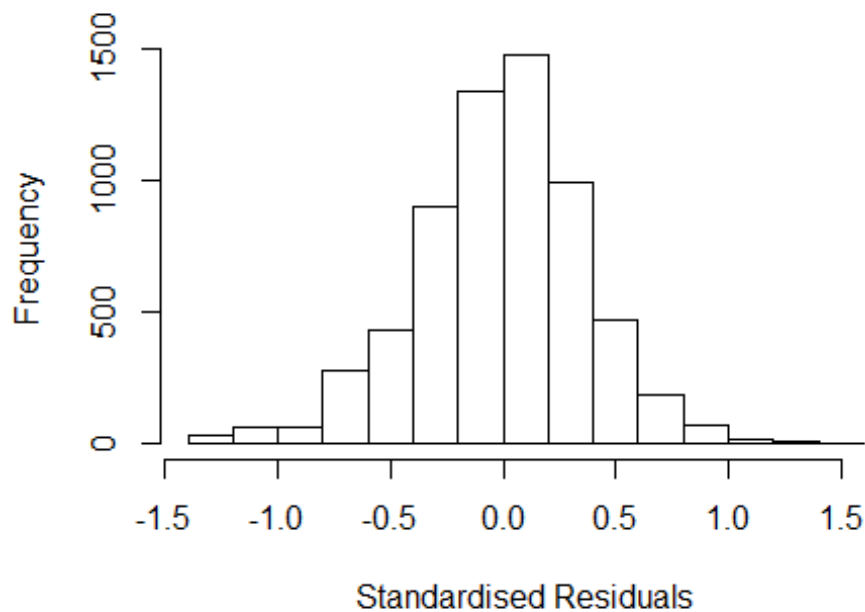


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 100, df = 1, p-value <2e-16
```



```
## [1] "Female first author team size 2018 geometric mean: 5.72221921982254"
## [1] "Male first author team size 2018 geometric mean: 4.2323488434449"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.64018258081074"
## [1] "Male last author team size 2018 geometric mean: 4.57869223040378"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1          1.013
## LastAuthorFemale  1.022 1          1.011
## UniqueAuthors    1.113 4          1.013
## Year              1.138 16         1.004
```

## Residuals from first and last author and team size



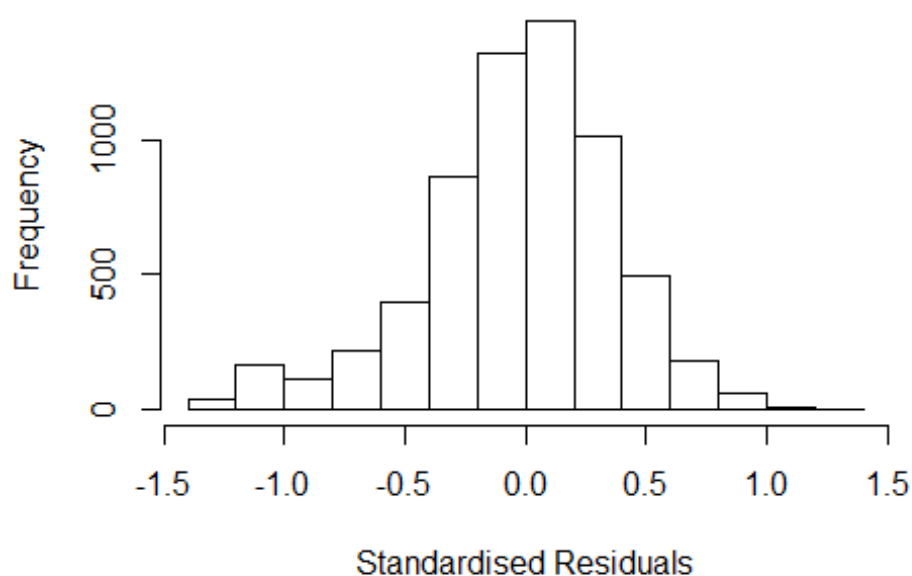
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.333 -0.235  0.009  0.226  1.457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.71e-01   4.60e-02  14.59  <2e-16 ***
## FirstAuthorFemale1 -1.37e-02   9.10e-03  -1.51   0.1312
## LastAuthorFemale1 -2.86e-04   1.09e-02  -0.03   0.9791
## UniqueAuthors2    4.24e-01   4.45e-02   9.52  <2e-16 ***
## UniqueAuthors3    4.77e-01   4.40e-02  10.84  <2e-16 ***
## UniqueAuthors4    5.05e-01   4.36e-02  11.59  <2e-16 ***
## UniqueAuthors5    6.13e-01   4.29e-02  14.28  <2e-16 ***
## Year1997          5.25e-02   3.04e-02   1.73   0.0842 .
## Year1998          4.98e-02   2.83e-02   1.76   0.0781 .
## Year1999          2.29e-02   2.78e-02   0.83   0.4093
```

```

## Year2000          7.29e-03   3.53e-02    0.21   0.8363
## Year2001          2.40e-02   3.23e-02    0.74   0.4564
## Year2002         -2.14e-02   2.59e-02   -0.83   0.4089
## Year2003         -8.21e-02   2.82e-02   -2.91   0.0036 **
## Year2004         -7.60e-02   2.69e-02   -2.82   0.0048 **
## Year2005         -7.40e-02   2.64e-02   -2.80   0.0051 **
## Year2006         -7.44e-02   2.66e-02   -2.79   0.0053 **
## Year2007         -6.91e-02   2.75e-02   -2.51   0.0121 *
## Year2008         -4.27e-02   2.74e-02   -1.56   0.1199
## Year2009         -5.66e-02   2.78e-02   -2.04   0.0416 *
## Year2010          6.13e-05   2.89e-02    0.00   0.9983
## Year2011         -4.00e-02   2.99e-02   -1.34   0.1809
## Year2012         -1.08e-02   2.86e-02   -0.38   0.7062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.335
## Multiple R-squared:  0.162, Adjusted R-squared:  0.159
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 532 weights are ~= 1. The remaining 5765 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0195 0.8610 0.9490 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.59e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1          1.015
## LastAuthorFemale 1.020 1          1.010
## Year          1.048 16          1.001

```

## Residuals from first and last author



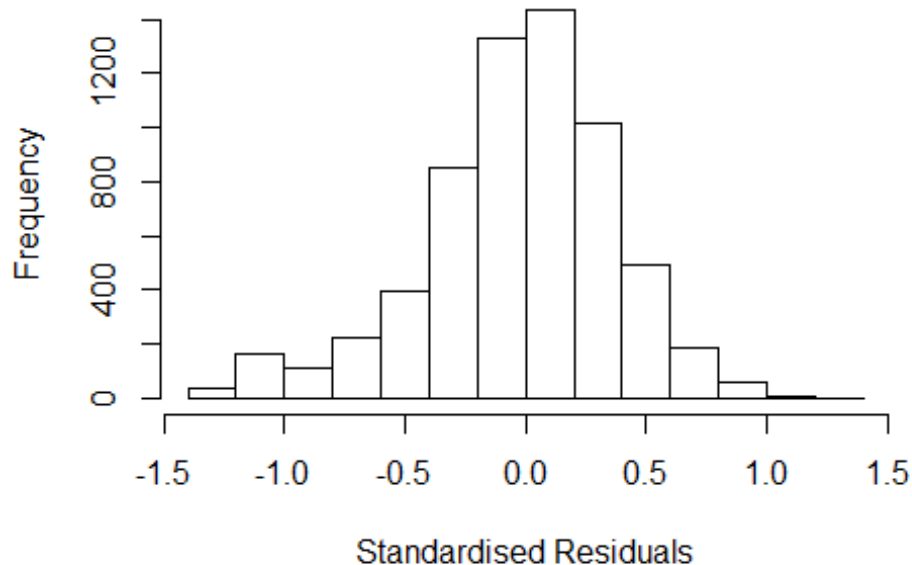
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22806 -0.24145  0.00555  0.22760  1.22905
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.14316    0.02091   54.66  <2e-16 ***
## FirstAuthorFemale1  0.00769    0.00959    0.80   0.4227
## LastAuthorFemale1 -0.00836    0.01106   -0.76   0.4499
## Year1997          0.06307    0.03005    2.10   0.0359 *
## Year1998          0.06727    0.02727    2.47   0.0136 *
## Year1999          0.03960    0.02754    1.44   0.1504
## Year2000          0.03815    0.03579    1.07   0.2866
## Year2001          0.07478    0.03216    2.33   0.0201 *
## Year2002          0.04262    0.02604    1.64   0.1018
## Year2003         -0.03593    0.02896   -1.24   0.2148
## Year2004         -0.03473    0.02756   -1.26   0.2077
## Year2005         -0.01118    0.02662   -0.42   0.6745
```

```

## Year2006      -0.02714    0.02725   -1.00    0.3194
## Year2007      -0.00943    0.02737   -0.34    0.7306
## Year2008       0.01555    0.02795    0.56    0.5779
## Year2009       0.00629    0.02827    0.22    0.8238
## Year2010       0.07721    0.02888    2.67    0.0075 **
## Year2011       0.03939    0.03114    1.26    0.2060
## Year2012       0.06984    0.02855    2.45    0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.0107, Adjusted R-squared:  0.00785
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 543 weights are ~= 1. The remaining 5754 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.187  0.862  0.951  0.883  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1      1.014
## Year              1.029 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22548 -0.24083 0.00592 0.22826 1.22277
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14194 0.02083 54.83 <2e-16 ***
## FirstAuthorFemale1 0.00740 0.00962 0.77 0.4419
## Year1997 0.06246 0.03004 2.08 0.0376 *
## Year1998 0.06742 0.02728 2.47 0.0135 *
## Year1999 0.03884 0.02754 1.41 0.1584
## Year2000 0.03729 0.03580 1.04 0.2977
## Year2001 0.07374 0.03216 2.29 0.0219 *
## Year2002 0.04194 0.02603 1.61 0.1072
## Year2003 -0.03631 0.02898 -1.25 0.2103
## Year2004 -0.03519 0.02756 -1.28 0.2017
## Year2005 -0.01197 0.02661 -0.45 0.6528
## Year2006 -0.02762 0.02726 -1.01 0.3110
```

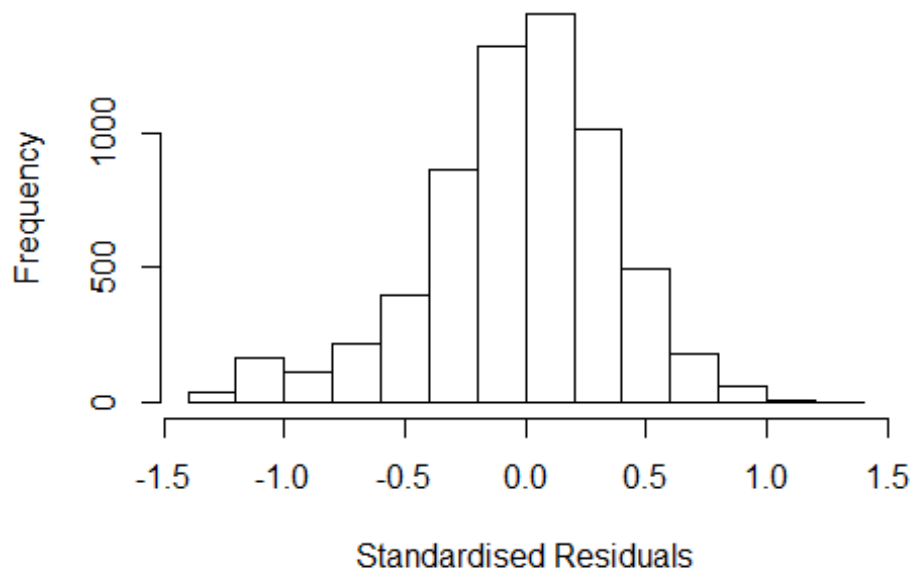
```

## Year2007          -0.01019    0.02736   -0.37    0.7097
## Year2008          0.01454    0.02796    0.52    0.6032
## Year2009          0.00548    0.02826    0.19    0.8463
## Year2010          0.07614    0.02887    2.64    0.0084 **
## Year2011          0.03849    0.03114    1.24    0.2164
## Year2012          0.06875    0.02851    2.41    0.0159 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.0106, Adjusted R-squared:  0.00792
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 549 weights are ~= 1. The remaining 5748 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.190  0.862  0.951  0.883  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.019 1          1.010
## Year            1.019 16          1.001

```



## Residuals from last author



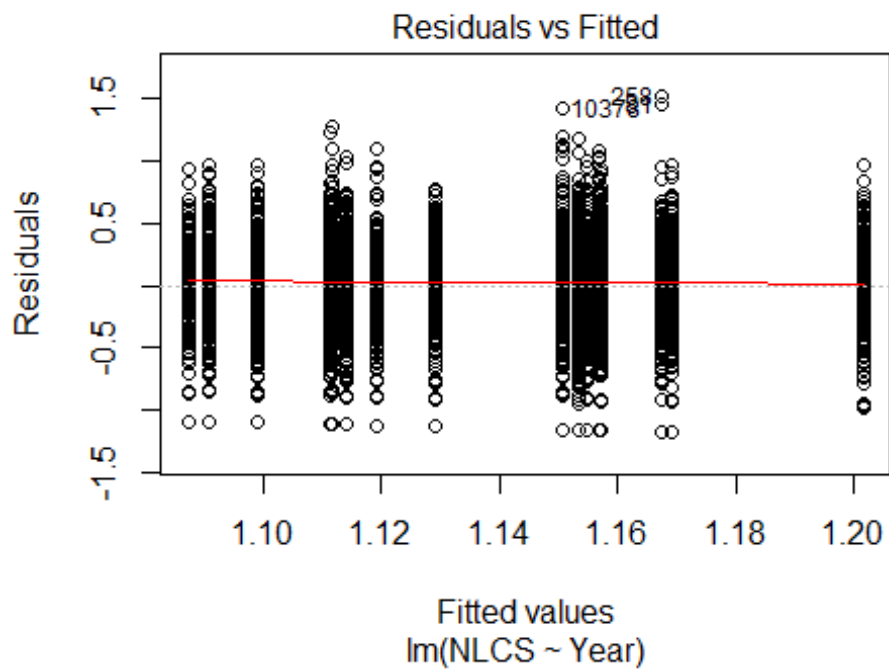
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22415 -0.24215  0.00585  0.22839  1.22537
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14604    0.02061   55.59  <2e-16 ***
## LastAuthorFemale1 -0.00795    0.01109   -0.72   0.4737
## Year1997         0.06326    0.03005    2.11   0.0353 *
## Year1998         0.06734    0.02726    2.47   0.0135 *
## Year1999         0.04007    0.02753    1.46   0.1456
## Year2000         0.03854    0.03573    1.08   0.2808
## Year2001         0.07506    0.03215    2.33   0.0196 *
## Year2002         0.04339    0.02598    1.67   0.0950 .
## Year2003        -0.03523    0.02892   -1.22   0.2233
## Year2004        -0.03395    0.02751   -1.23   0.2172
## Year2005        -0.01018    0.02651   -0.38   0.7010
## Year2006        -0.02632    0.02718   -0.97   0.3331
```

```

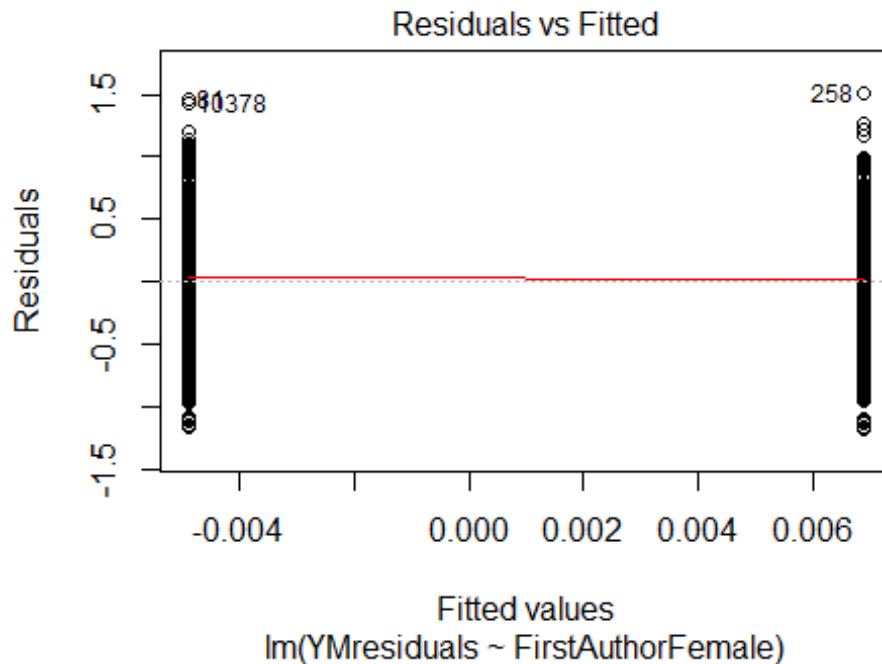
## Year2007          -0.00869      0.02733    -0.32    0.7505
## Year2008           0.01637      0.02790      0.59    0.5573
## Year2009           0.00757      0.02817      0.27    0.7881
## Year2010           0.07811      0.02883      2.71    0.0068 **
## Year2011           0.04045      0.03105      1.30    0.1927
## Year2012           0.07045      0.02852      2.47    0.0135 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.0106, Adjusted R-squared:  0.0079
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 539 weights are ~= 1. The remaining 5758 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.190  0.862  0.951   0.883  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.59e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6297"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2404"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   603  536  558  529  582  617  585  450  472  498  484  486  461  376  397
## 2011 2012
##   394  375
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   281  259  264  294  217  168  378  294  306  342  319  321  352  290  300
## 2011 2012

```

```
## 283 269
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 242 233 241 254 194 152 324 258 271 296 276 286 311 253 270
## 2011 2012
## 249 244
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 100, df = 16, p-value = 3e-14
```

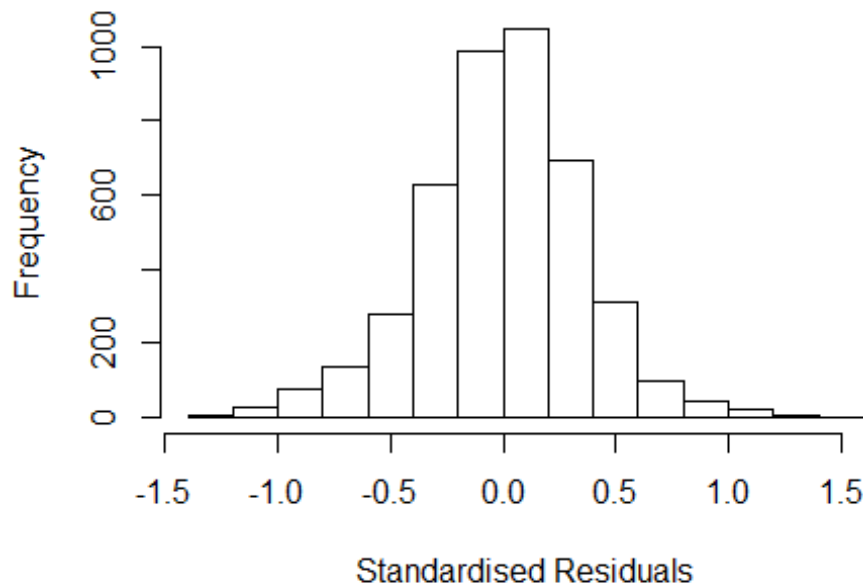


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 19, df = 1, p-value = 1e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.41571559060901"
## [1] "Male first author team size 2018 geometric mean: 4.33623372081236"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6300, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.12250941693524"
## [1] "Male last author team size 2018 geometric mean: 4.51297514500742"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5300, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1          1.022
## LastAuthorFemale  1.032 1          1.016
## UniqueAuthors     1.198 4          1.023
## Year              1.269 16          1.007
```

## Residuals from first and last author and team size



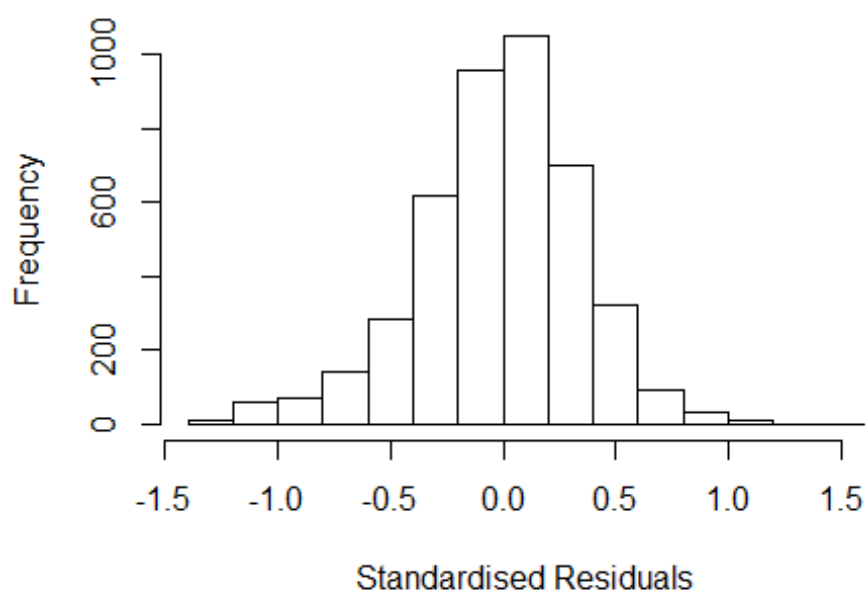
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25687 -0.21781 0.00678 0.22005 1.41204
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9009 0.0511 17.63 < 2e-16 ***
## FirstAuthorFemale1 -0.0128 0.0106 -1.21 0.2263
## LastAuthorFemale1 -0.0357 0.0149 -2.39 0.0168 *
## UniqueAuthors2 0.2360 0.0498 4.74 2.2e-06 ***
## UniqueAuthors3 0.2680 0.0490 5.46 4.9e-08 ***
## UniqueAuthors4 0.3161 0.0489 6.46 1.2e-10 ***
## UniqueAuthors5 0.3809 0.0487 7.83 6.3e-15 ***
## Year1997 -0.0153 0.0321 -0.48 0.6322
## Year1998 0.0266 0.0307 0.87 0.3859
## Year1999 -0.0249 0.0311 -0.80 0.4231
```

```

## Year2000          -0.0104      0.0327   -0.32   0.7498
## Year2001          -0.0711      0.0378   -1.88   0.0600 .
## Year2002          -0.0189      0.0291   -0.65   0.5175
## Year2003          -0.0342      0.0312   -1.10   0.2723
## Year2004          -0.0579      0.0306   -1.89   0.0583 .
## Year2005          -0.0661      0.0281   -2.35   0.0189 *
## Year2006          -0.0818      0.0287   -2.85   0.0044 **
## Year2007          -0.0573      0.0308   -1.86   0.0629 .
## Year2008          -0.0536      0.0294   -1.82   0.0681 .
## Year2009          -0.0332      0.0295   -1.13   0.2604
## Year2010          -0.0540      0.0332   -1.63   0.1040
## Year2011          -0.0279      0.0343   -0.81   0.4158
## Year2012          -0.0155      0.0345   -0.45   0.6541
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.321
## Multiple R-squared:  0.0762, Adjusted R-squared:  0.0715
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 395 weights are ~= 1. The remaining 3959 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0136 0.8610 0.9490 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.30e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.045 1 1.022
## LastAuthorFemale 1.029 1 1.014
## Year 1.075 16 1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21585 -0.22448  0.00512  0.21803  1.50257
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18017    0.02208   53.45  <2e-16 ***
## FirstAuthorFemale1 -0.00173    0.01077   -0.16   0.8720
## LastAuthorFemale1 -0.04287    0.01549   -2.77   0.0057 **
## Year1997        -0.01540    0.03212   -0.48   0.6315
## Year1998         0.03628    0.03048    1.19   0.2340
## Year1999        -0.01422    0.03045   -0.47   0.6405
## Year2000         0.00294    0.03285    0.09   0.9286
## Year2001        -0.05706    0.03728   -1.53   0.1260
## Year2002         0.01779    0.02882    0.62   0.5370
## Year2003        -0.01843    0.03107   -0.59   0.5531
## Year2004        -0.04166    0.03038   -1.37   0.1704
## Year2005        -0.04002    0.02851   -1.40   0.1604
```

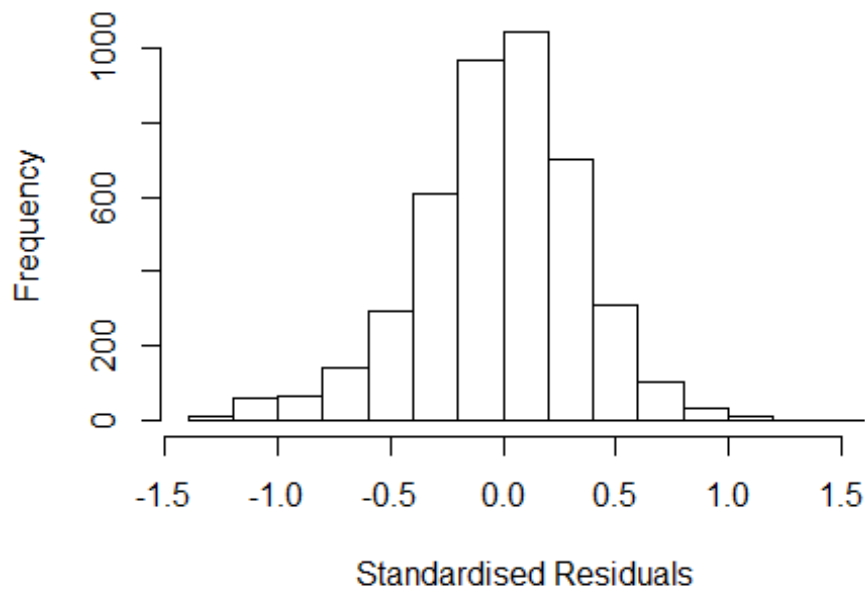
```

## Year2006      -0.06047    0.02935   -2.06    0.0394 *
## Year2007      -0.02478    0.03110   -0.80    0.4256
## Year2008      -0.02465    0.02937   -0.84    0.4014
## Year2009       0.00390    0.02964    0.13    0.8954
## Year2010       0.00766    0.03285    0.23    0.8156
## Year2011       0.01686    0.03437    0.49    0.6238
## Year2012       0.03568    0.03395    1.05    0.2933
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.00854,    Adjusted R-squared:  0.00442
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 379 weights are ~= 1. The remaining 3975 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0017 0.8640 0.9500 0.8880 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.045 1      1.022
## Year      1.045 16      1.001

```



## Residuals from first author



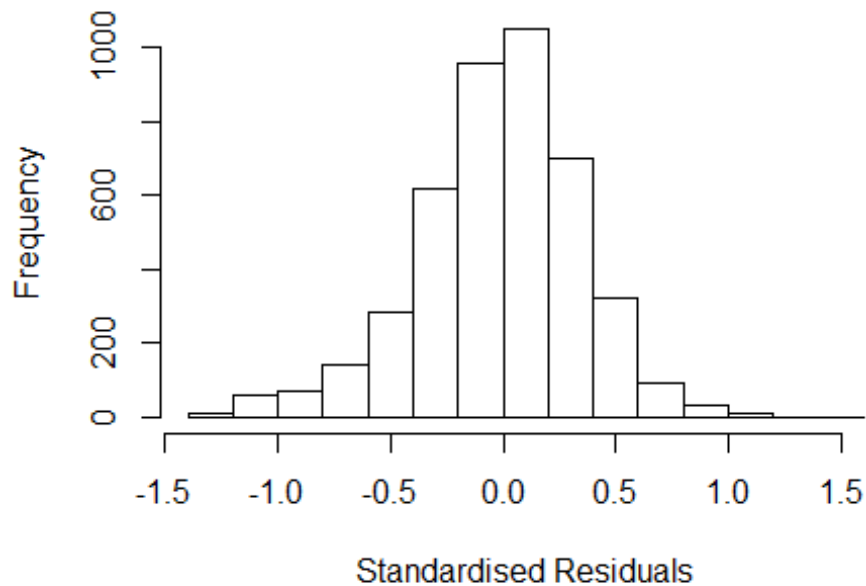
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20782 -0.22250  0.00571  0.21798  1.50846
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.175317   0.022028   53.36  <2e-16 ***
## FirstAuthorFemale1 -0.002781   0.010790   -0.26   0.797
## Year1997        -0.017326   0.032036   -0.54   0.589
## Year1998         0.035816   0.030457    1.18   0.240
## Year1999        -0.014523   0.030400   -0.48   0.633
## Year2000         0.000303   0.032868    0.01   0.993
## Year2001        -0.056822   0.037103   -1.53   0.126
## Year2002         0.016394   0.028838    0.57   0.570
## Year2003        -0.018695   0.030971   -0.60   0.546
## Year2004        -0.041460   0.030384   -1.36   0.172
## Year2005        -0.040082   0.028518   -1.41   0.160
## Year2006        -0.060034   0.029320   -2.05   0.041 *
```

```

## Year2007          -0.028781    0.030963    -0.93    0.353
## Year2008          -0.027238    0.029326    -0.93    0.353
## Year2009           0.002132    0.029723     0.07    0.943
## Year2010           0.006471    0.032717     0.20    0.843
## Year2011           0.014497    0.034421     0.42    0.674
## Year2012           0.032506    0.033943     0.96    0.338
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.00652,    Adjusted R-squared:  0.00262
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 368 weights are ~= 1. The remaining 3986 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0011 0.8620 0.9510 0.8880 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1      1.015
## Year             1.029 16      1.001

```

## Residuals from last author



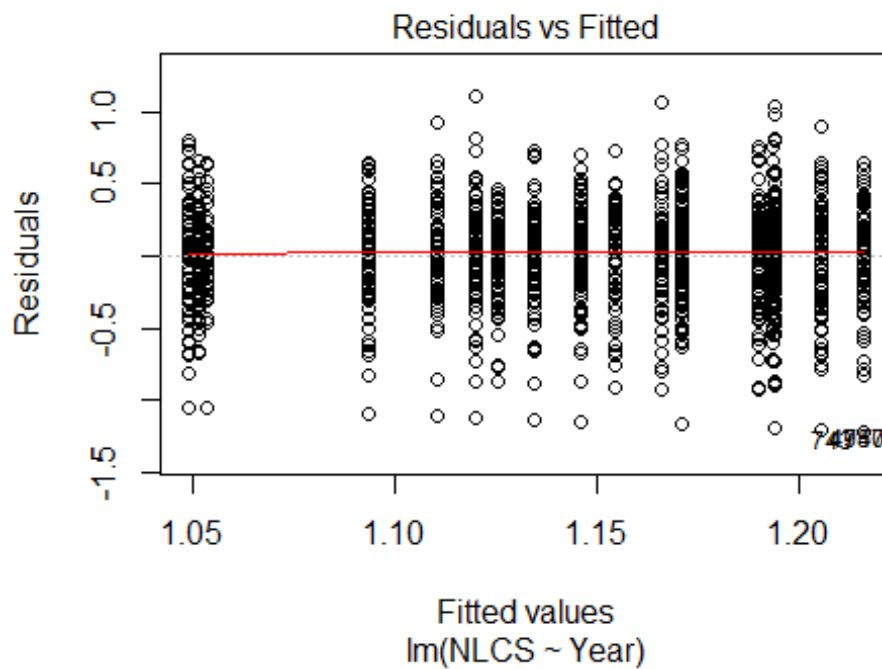
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21509 -0.22462  0.00571  0.21791  1.50141
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17959    0.02189   53.90  <2e-16 ***
## LastAuthorFemale1 -0.04294    0.01551   -2.77   0.0057 **
## Year1997      -0.01546    0.03211   -0.48   0.6302
## Year1998       0.03623    0.03048    1.19   0.2346
## Year1999      -0.01430    0.03044   -0.47   0.6386
## Year2000       0.00292    0.03285    0.09   0.9292
## Year2001      -0.05721    0.03727   -1.53   0.1249
## Year2002       0.01772    0.02881    0.62   0.5384
## Year2003      -0.01865    0.03097   -0.60   0.5470
## Year2004      -0.04178    0.03037   -1.38   0.1690
## Year2005      -0.04026    0.02840   -1.42   0.1565
## Year2006      -0.06059    0.02932   -2.07   0.0388 *
```

```

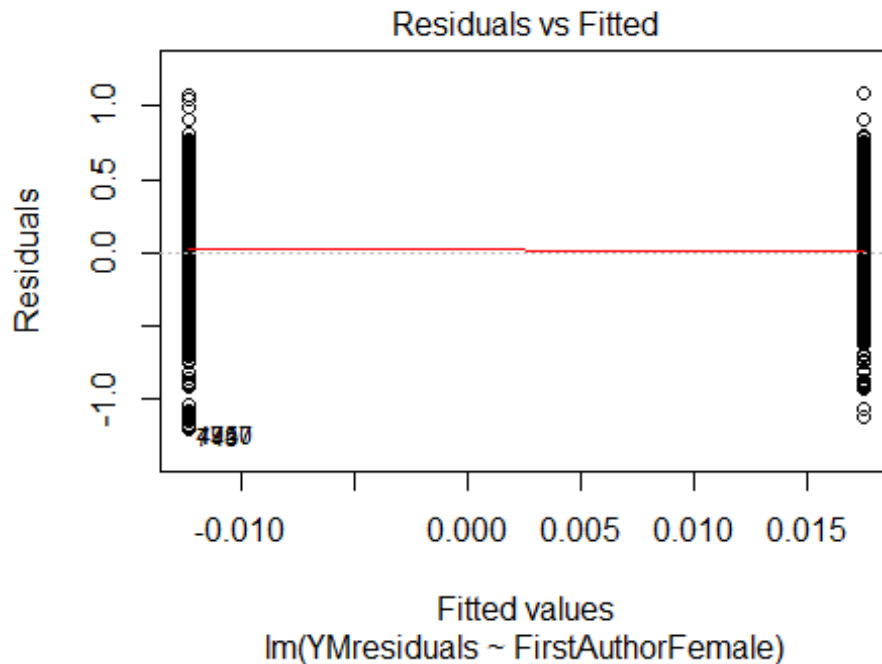
## Year2007          -0.02500      0.03101    -0.81    0.4201
## Year2008          -0.02496      0.02921    -0.85    0.3928
## Year2009           0.00362      0.02950     0.12    0.9025
## Year2010           0.00742      0.03276     0.23    0.8207
## Year2011           0.01660      0.03423     0.48    0.6278
## Year2012           0.03550      0.03387     1.05    0.2946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.00852,    Adjusted R-squared:  0.00464
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 375 weights are ~= 1. The remaining 3979 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8640 0.9500 0.8880 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4354"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2405"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 239 227 213 213 210 238 244 169 152 180 175 168 166 172 186
## 2011 2012
## 143 182
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 88 75 83 71 75 56 124 91 71 90 85 83 88 102 105
## 2011 2012

```

```
## 93 120
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 73 68 78 64 70 53 99 81 66 81 77 75 81 93 96
## 2011 2012
## 82 111
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```

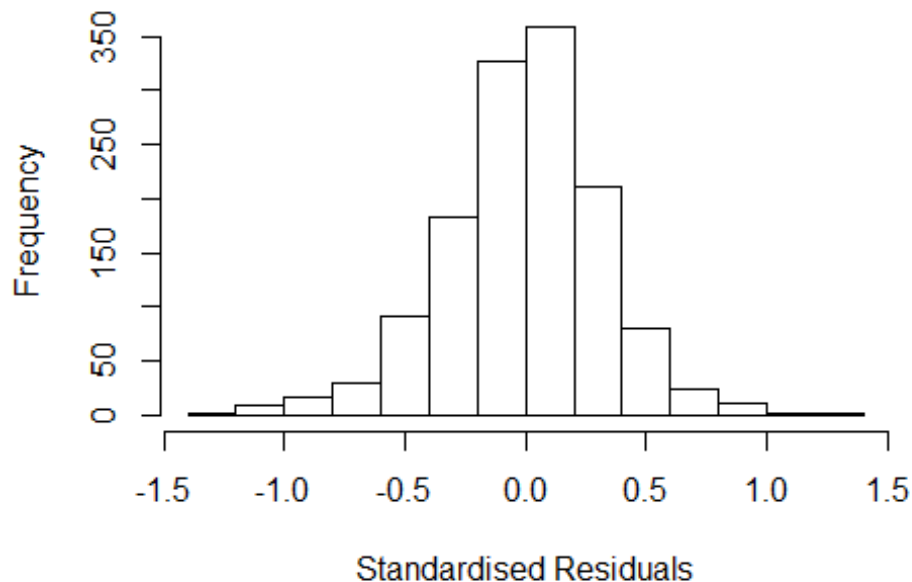


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.8, df = 1, p-value = 0.005
```



```
## [1] "Female first author team size 2018 geometric mean: 5.46424398528896"
## [1] "Male first author team size 2018 geometric mean: 4.44495510760367"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.66435906602096"
## [1] "Male last author team size 2018 geometric mean: 4.65747172876145"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.092 1          1.045
## LastAuthorFemale  1.086 1          1.042
## UniqueAuthors     1.456 4          1.048
## Year              1.560 16         1.014
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2639 -0.1952 0.0137 0.1962 1.2899
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06209 0.05942 17.87 < 2e-16 ***
## FirstAuthorFemale1 0.01366 0.01763 0.77 0.4387
## LastAuthorFemale1 -0.03261 0.02071 -1.58 0.1155
## UniqueAuthors2 0.06511 0.05098 1.28 0.2018
## UniqueAuthors3 0.04379 0.05032 0.87 0.3843
## UniqueAuthors4 0.14728 0.04900 3.01 0.0027 **
## UniqueAuthors5 0.25798 0.04703 5.49 4.9e-08 ***
## Year1997 0.00649 0.06050 0.11 0.9145
## Year1998 0.07187 0.04834 1.49 0.1373
## Year1999 -0.06984 0.05371 -1.30 0.1937
```

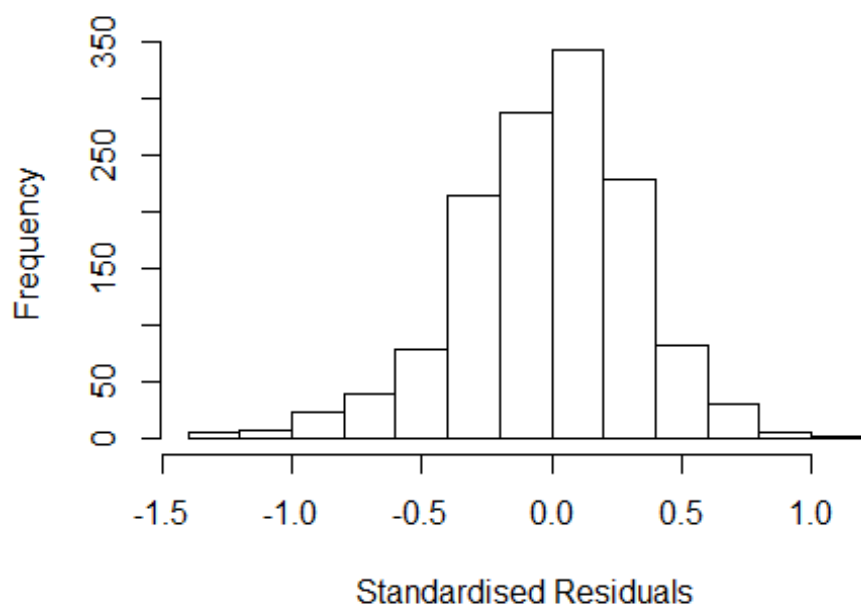
```

## Year2000      -0.07906      0.05271      -1.50      0.1339
## Year2001      -0.11963      0.05679      -2.11      0.0353 *
## Year2002      -0.02722      0.04934      -0.55      0.5813
## Year2003      -0.13451      0.04731      -2.84      0.0045 **
## Year2004      -0.13622      0.06129      -2.22      0.0264 *
## Year2005      -0.08360      0.04643      -1.80      0.0720 .
## Year2006      -0.10899      0.04913      -2.22      0.0267 *
## Year2007      -0.02819      0.04858      -0.58      0.5619
## Year2008      -0.02962      0.04992      -0.59      0.5531
## Year2009      -0.06451      0.04873      -1.32      0.1858
## Year2010      -0.03064      0.04866      -0.63      0.5290
## Year2011       0.01977      0.05800       0.34      0.7332
## Year2012       0.00293      0.04668       0.06      0.9500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.295
## Multiple R-squared:  0.127, Adjusted R-squared:  0.112
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1240 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0162 0.8640 0.9520 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.42e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.071 1      1.035
## LastAuthorFemale  1.049 1      1.024
## Year              1.124 16      1.004

```



## Residuals from first and last author



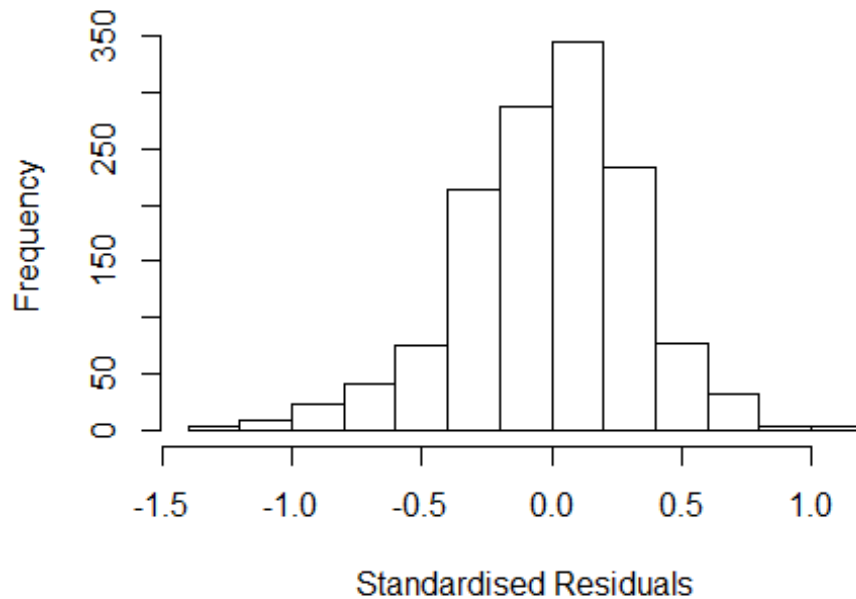
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2687 -0.2164  0.0112  0.2093  1.1382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.164430   0.037172   31.33  <2e-16 ***
## FirstAuthorFemale1  0.017308   0.018459    0.94   0.349
## LastAuthorFemale1 -0.054118   0.022252   -2.43   0.015 *
## Year1997          0.016992   0.060584    0.28   0.779
## Year1998          0.101149   0.051578    1.96   0.050 .
## Year1999         -0.046438   0.057490   -0.81   0.419
## Year2000         -0.045129   0.056295   -0.80   0.423
## Year2001         -0.106449   0.057361   -1.86   0.064 .
## Year2002          0.038603   0.050659    0.76   0.446
## Year2003         -0.101414   0.050087   -2.02   0.043 *
## Year2004         -0.098163   0.065967   -1.49   0.137
## Year2005         -0.012517   0.046598   -0.27   0.788
```

```

## Year2006      -0.041836    0.052158   -0.80    0.423
## Year2007      0.041383    0.050653    0.82    0.414
## Year2008      0.029676    0.051045    0.58    0.561
## Year2009      0.000236    0.049565    0.00    0.996
## Year2010      0.036237    0.050802    0.71    0.476
## Year2011      0.104251    0.055269    1.89    0.059 .
## Year2012      0.069777    0.049313    1.41    0.157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.318
## Multiple R-squared:  0.0387, Adjusted R-squared:  0.0256
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 111 weights are ~= 1. The remaining 1237 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0746 0.8760 0.9520 0.8960 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.072 1      1.035
## Year      1.072 16      1.002

```

## Residuals from first author



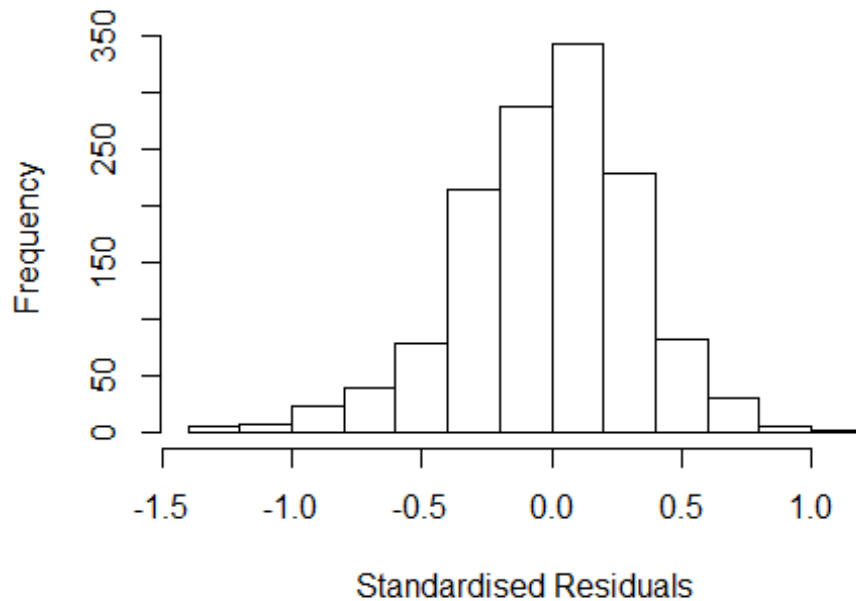
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2578 -0.2233 0.0114 0.2076 1.0950
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15463 0.03766 30.66 <2e-16 ***
## FirstAuthorFemale1 0.01349 0.01861 0.73 0.469
## Year1997 0.01660 0.06151 0.27 0.787
## Year1998 0.09607 0.05167 1.86 0.063 .
## Year1999 -0.04307 0.05839 -0.74 0.461
## Year2000 -0.04950 0.05711 -0.87 0.386
## Year2001 -0.10585 0.05822 -1.82 0.069 .
## Year2002 0.03616 0.05121 0.71 0.480
## Year2003 -0.10041 0.05077 -1.98 0.048 *
## Year2004 -0.09803 0.06652 -1.47 0.141
## Year2005 -0.01183 0.04741 -0.25 0.803
## Year2006 -0.03910 0.05271 -0.74 0.458
```

```

## Year2007          0.03689    0.05066    0.73    0.467
## Year2008          0.02369    0.05176    0.46    0.647
## Year2009         -0.00025    0.05008    0.00    0.996
## Year2010          0.03084    0.05074    0.61    0.543
## Year2011          0.10318    0.05532    1.87    0.062 .
## Year2012          0.07077    0.04989    1.42    0.156
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.319
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.0215
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 121 weights are ~= 1. The remaining 1227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.084  0.876  0.950  0.895  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.049 1          1.024
## Year            1.049 16          1.001

```

## Residuals from last author



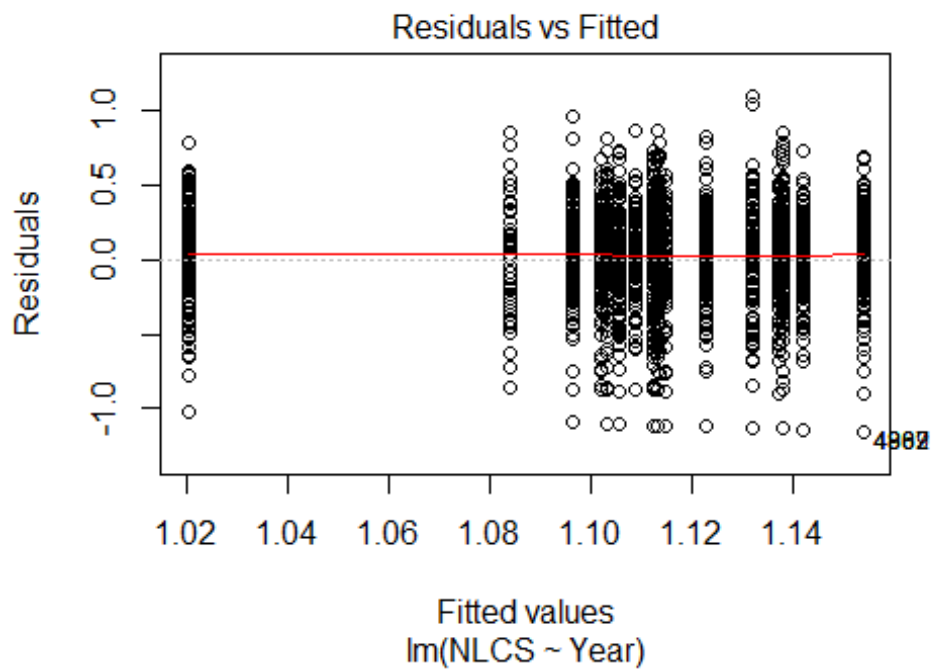
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2774 -0.2199 0.0103 0.2052 1.1458
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16943 0.03688 31.70 <2e-16 ***
## LastAuthorFemale1 -0.05244 0.02229 -2.35 0.019 *
## Year1997 0.01841 0.06048 0.30 0.761
## Year1998 0.10136 0.05161 1.96 0.050 *
## Year1999 -0.04568 0.05745 -0.80 0.427
## Year2000 -0.04444 0.05645 -0.79 0.431
## Year2001 -0.10598 0.05744 -1.84 0.065 .
## Year2002 0.04180 0.05049 0.83 0.408
## Year2003 -0.09797 0.04991 -1.96 0.050 *
## Year2004 -0.09596 0.06567 -1.46 0.144
## Year2005 -0.00878 0.04625 -0.19 0.849
## Year2006 -0.03875 0.05189 -0.75 0.455
```

```

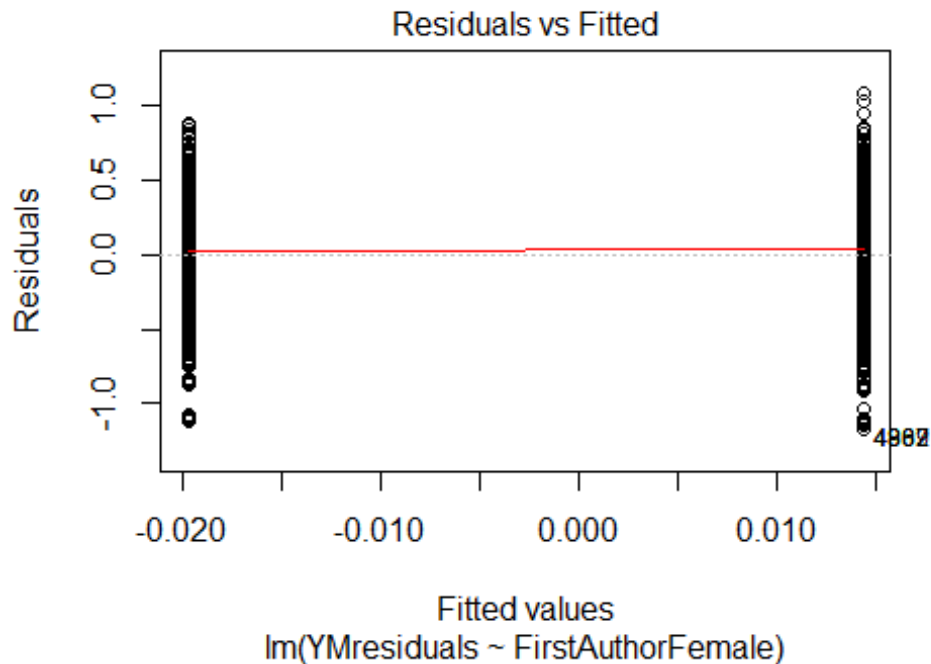
## Year2007      0.04279      0.05050      0.85      0.397
## Year2008      0.03243      0.05068      0.64      0.522
## Year2009      0.00242      0.04921      0.05      0.961
## Year2010      0.03759      0.05074      0.74      0.459
## Year2011      0.10794      0.05477      1.97      0.049 *
## Year2012      0.07056      0.04936      1.43      0.153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.317
## Multiple R-squared:  0.0383, Adjusted R-squared:  0.026
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1235 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0687 0.8770 0.9520 0.8950 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.42e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1348"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2406"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 229 223 218 222 201 243 242 153 203 239 237 269 229 240 201
## 2011 2012
## 223 207
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 98 100 120 91 48 130 102 145 164 143 160 150 154 128
## 2011 2012

```

```
## 145 146
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 85 90 105 77 43 115 92 134 147 129 147 135 137 119
## 2011 2012
## 131 134
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```



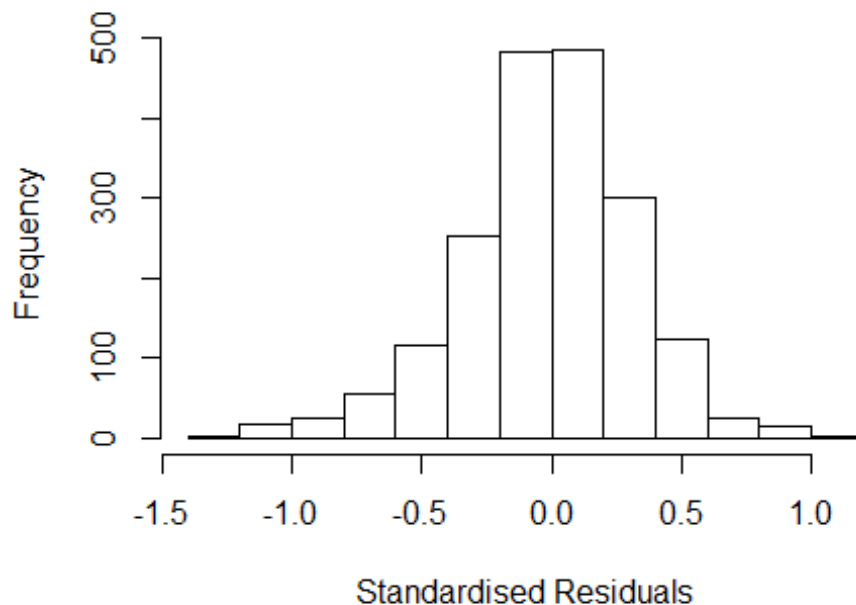
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.2, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 5.9196815973974"
## [1] "Male first author team size 2018 geometric mean: 6.04989885561656"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 7.09247086197954"
## [1] "Male last author team size 2018 geometric mean: 5.47147700212433"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.076 1      1.038
## LastAuthorFemale  1.059 1      1.029
## UniqueAuthors    1.242 4      1.027
## Year             1.335 16      1.009
```



## Residuals from first and last author and team size



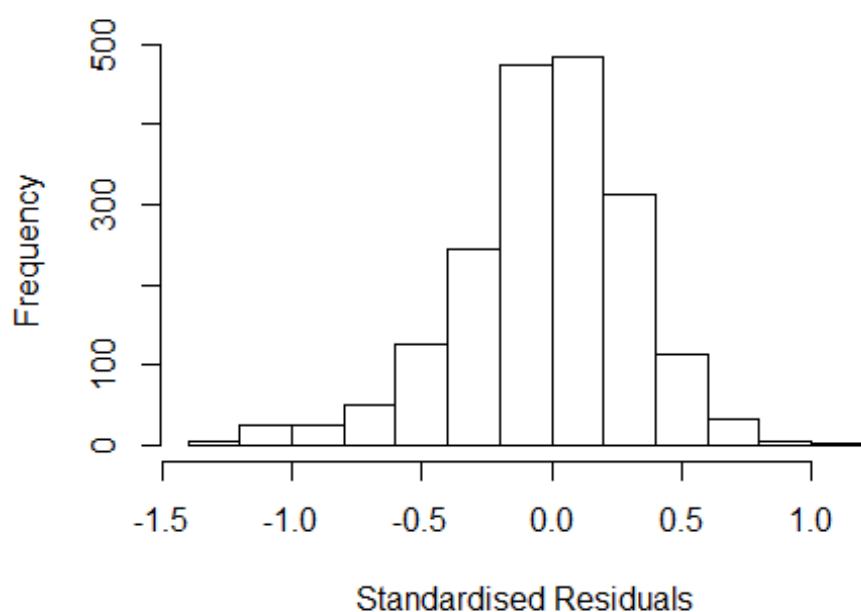
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.244692 -0.196800 -0.000255 0.190324 1.011335
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97000 0.07339 13.22 < 2e-16 ***
## FirstAuthorFemale1 -0.04799 0.01503 -3.19 0.00143 **
## LastAuthorFemale1 -0.06394 0.01876 -3.41 0.00067 ***
## UniqueAuthors2 0.18901 0.06887 2.74 0.00612 **
## UniqueAuthors3 0.17338 0.06756 2.57 0.01035 *
## UniqueAuthors4 0.20239 0.06754 3.00 0.00277 **
## UniqueAuthors5 0.27097 0.06682 4.05 5.2e-05 ***
## Year1997 0.00529 0.04748 0.11 0.91137
## Year1998 -0.02909 0.04684 -0.62 0.53466
## Year1999 0.00525 0.04568 0.11 0.90854
```

```

## Year2000      -0.05359    0.04665   -1.15  0.25078
## Year2001      -0.05781    0.08354   -0.69  0.48900
## Year2002      -0.03390    0.04619   -0.73  0.46309
## Year2003      -0.04956    0.04862   -1.02  0.30815
## Year2004      -0.03433    0.04295   -0.80  0.42426
## Year2005      -0.09330    0.04496   -2.08  0.03812 *
## Year2006      -0.03134    0.04383   -0.72  0.47468
## Year2007      -0.01614    0.04419   -0.37  0.71499
## Year2008      -0.02718    0.04701   -0.58  0.56324
## Year2009      -0.01600    0.04618   -0.35  0.72908
## Year2010       0.00455    0.04740    0.10  0.92355
## Year2011       0.02697    0.04434    0.61  0.54303
## Year2012       0.00372    0.04671    0.08  0.93649
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.296
## Multiple R-squared:  0.0645, Adjusted R-squared:  0.0535
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1743 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0374 0.8600 0.9520 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.26e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.068 1          1.033
## LastAuthorFemale 1.037 1          1.018
## Year              1.097 16          1.003

```

## Residuals from first and last author



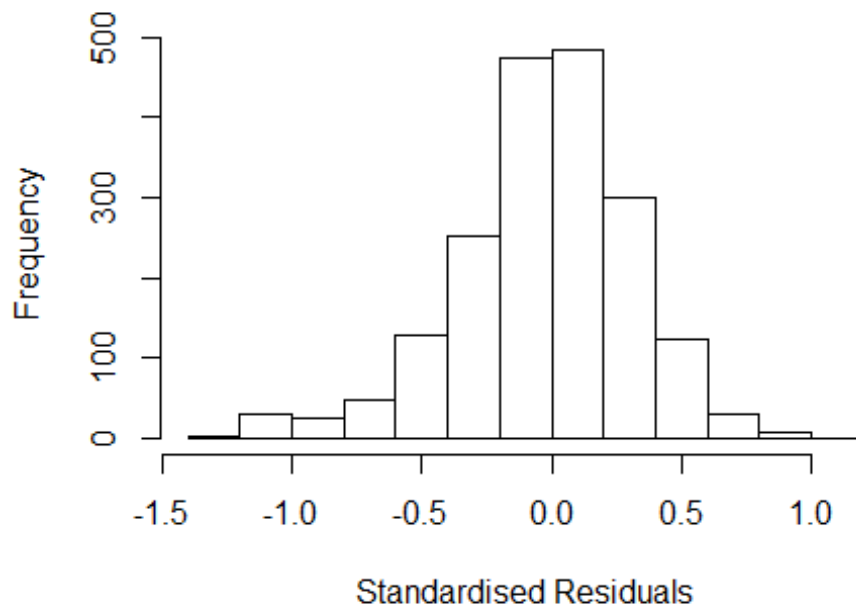
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21648 -0.20185 -0.00136 0.19584 1.02964
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16903 0.03524 33.18 < 2e-16 ***
## FirstAuthorFemale1 -0.04495 0.01522 -2.95 0.00318 **
## LastAuthorFemale1 -0.07143 0.01900 -3.76 0.00017 ***
## Year1997 0.00967 0.04713 0.21 0.83739
## Year1998 -0.02476 0.04630 -0.53 0.59283
## Year1999 0.02392 0.04444 0.54 0.59044
## Year2000 -0.03440 0.04583 -0.75 0.45296
## Year2001 -0.03587 0.08016 -0.45 0.65456
## Year2002 -0.00234 0.04560 -0.05 0.95901
## Year2003 -0.01744 0.04822 -0.36 0.71758
## Year2004 -0.02174 0.04229 -0.51 0.60724
## Year2005 -0.07895 0.04554 -1.73 0.08316 .
```

```

## Year2006      -0.00794    0.04321   -0.18  0.85418
## Year2007      -0.00423    0.04346   -0.10  0.92247
## Year2008       0.00437    0.04580    0.10  0.92404
## Year2009       0.01339    0.04546    0.29  0.76830
## Year2010       0.02547    0.04764    0.53  0.59297
## Year2011       0.04745    0.04365    1.09  0.27719
## Year2012       0.04276    0.04574    0.93  0.34995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.301
## Multiple R-squared:  0.025, Adjusted R-squared:  0.0157
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 173 weights are ~= 1. The remaining 1728 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0654 0.8630 0.9510 0.8870 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1      1.031
## Year              1.062 16      1.002

```

## Residuals from first author



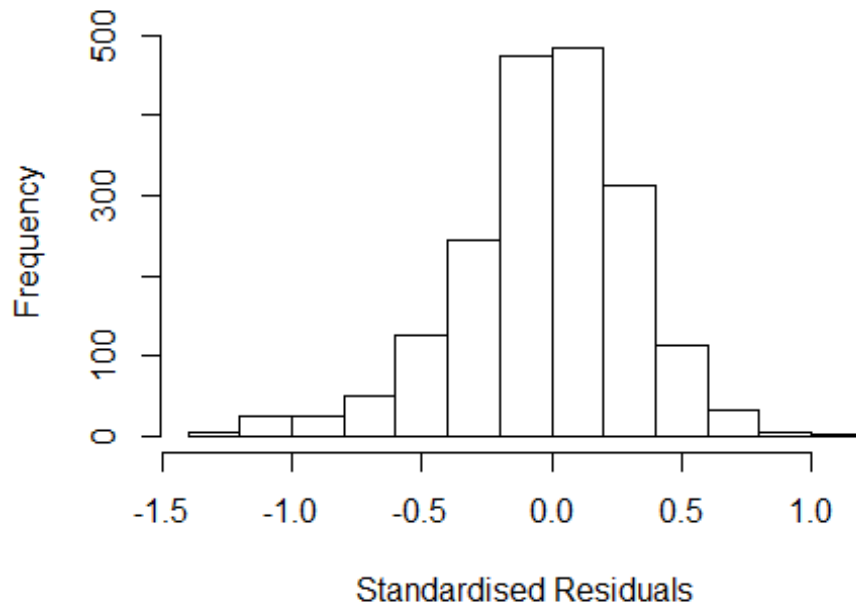
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20719 -0.20353 -0.00176 0.19580 1.03236
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16040 0.03512 33.04 <2e-16 ***
## FirstAuthorFemale1 -0.05016 0.01526 -3.29 0.001 **
## Year1997 0.00595 0.04748 0.13 0.900
## Year1998 -0.02771 0.04656 -0.60 0.552
## Year1999 0.02546 0.04442 0.57 0.567
## Year2000 -0.03485 0.04592 -0.76 0.448
## Year2001 -0.04724 0.07923 -0.60 0.551
## Year2002 -0.00432 0.04565 -0.09 0.925
## Year2003 -0.01758 0.04803 -0.37 0.714
## Year2004 -0.02448 0.04233 -0.58 0.563
## Year2005 -0.08355 0.04556 -1.83 0.067 .
## Year2006 -0.01154 0.04309 -0.27 0.789
```

```

## Year2007          -0.00976    0.04345   -0.22    0.822
## Year2008          0.00190    0.04581    0.04    0.967
## Year2009          0.00784    0.04556    0.17    0.863
## Year2010          0.02343    0.04789    0.49    0.625
## Year2011          0.04678    0.04397    1.06    0.287
## Year2012          0.03624    0.04590    0.79    0.430
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.301
## Multiple R-squared:  0.0166, Adjusted R-squared:  0.0077
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0726 0.8610 0.9510 0.8880 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.034 1      1.017
## Year      1.034 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19444 -0.20304 -0.00114  0.19716  1.05219
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.15332    0.03463   33.30 < 2e-16 ***
## LastAuthorFemale1 -0.07664    0.01903   -4.03 5.9e-05 ***
## Year1997         0.01120    0.04748    0.24  0.814
## Year1998        -0.02237    0.04643   -0.48  0.630
## Year1999         0.01967    0.04463    0.44  0.659
## Year2000        -0.03428    0.04620   -0.74  0.458
## Year2001        -0.03732    0.07952   -0.47  0.639
## Year2002        -0.00119    0.04557   -0.03  0.979
## Year2003        -0.01810    0.04831   -0.37  0.708
## Year2004        -0.02418    0.04224   -0.57  0.567
## Year2005        -0.08475    0.04547   -1.86  0.063 .
## Year2006        -0.01119    0.04323   -0.26  0.796
```

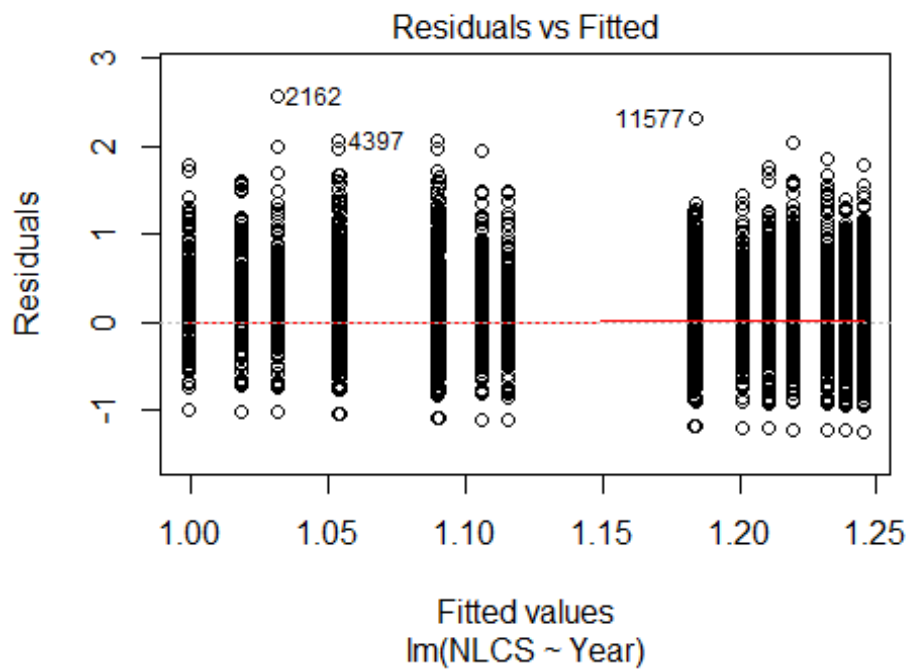
```

## Year2007          -0.00832      0.04315    -0.19      0.847
## Year2008           0.00173      0.04584      0.04      0.970
## Year2009           0.01186      0.04542      0.26      0.794
## Year2010           0.02116      0.04772      0.44      0.658
## Year2011           0.04094      0.04345      0.94      0.346
## Year2012           0.04112      0.04590      0.90      0.370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.302
## Multiple R-squared:  0.02,   Adjusted R-squared:  0.0112
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0822 0.8660 0.9510 0.8880 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1901"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2500"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1115 1097 1033 1037 1134 1021 885 828 938 825 955 892 1011 1037 1009
## 2011 2012
## 967 963
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 203 198 203 206 214 175 202 270 287 286 346 354 377 414 417
## 2011 2012

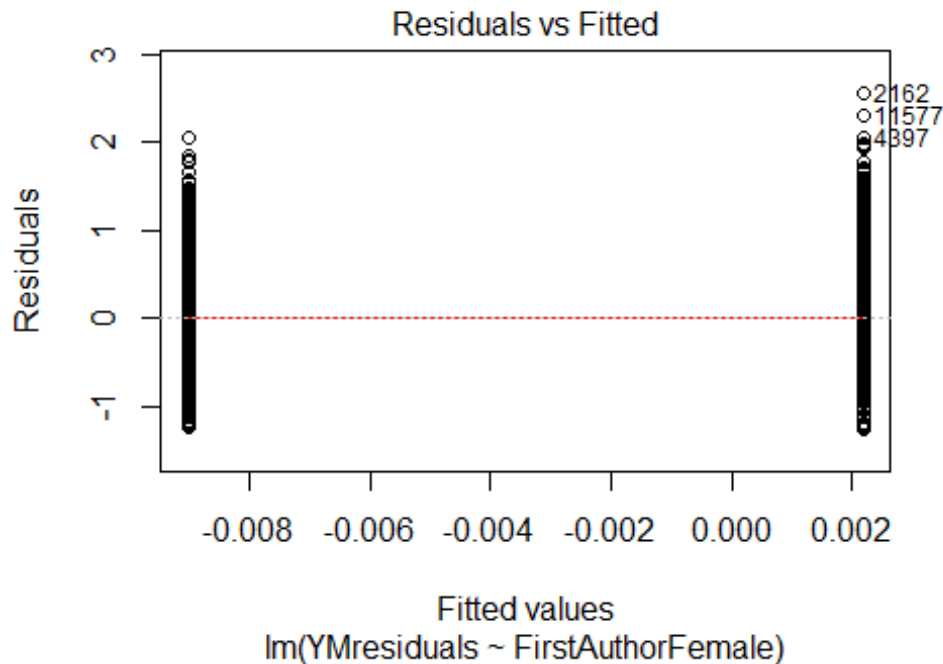
```



```
## 415 451
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 182 175 187 181 194 154 167 251 250 237 297 307 313 354 370
## 2011 2012
## 349 393
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 81, df = 16, p-value = 1e-10
```

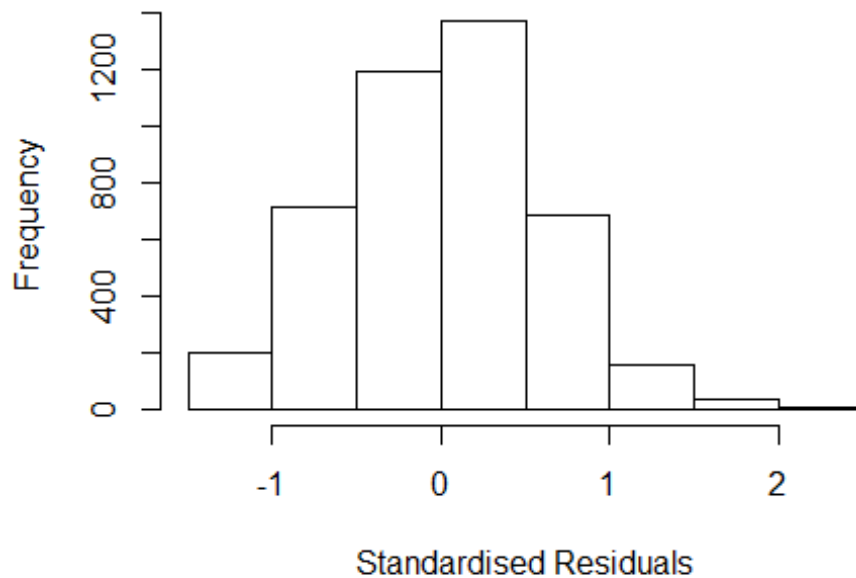


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.13, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 4.14983508429971"
## [1] "Male first author team size 2018 geometric mean: 3.58002231211064"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 39000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.93001139854316"
## [1] "Male last author team size 2018 geometric mean: 3.66094737995846"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.070 1          1.034
## LastAuthorFemale  1.026 1          1.013
## UniqueAuthors    1.172 4          1.020
## Year              1.174 16         1.005
```

## Residuals from first and last author and team size



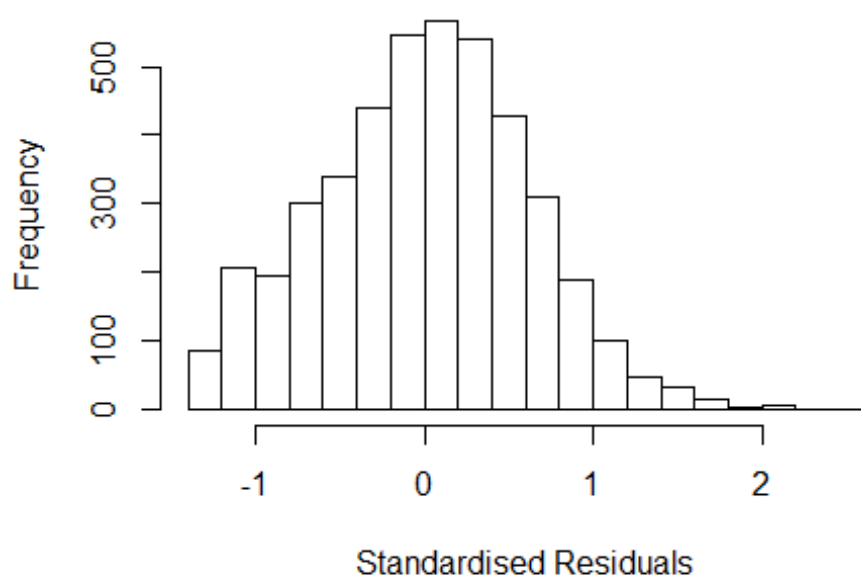
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4563 -0.4110 0.0234 0.4017 2.4670
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8800 0.0549 16.04 <2e-16 ***
## FirstAuthorFemale1 -0.0538 0.0239 -2.25 0.024 *
## LastAuthorFemale1 -0.0363 0.0335 -1.08 0.279
## UniqueAuthors2 0.2790 0.0320 8.71 <2e-16 ***
## UniqueAuthors3 0.3008 0.0313 9.60 <2e-16 ***
## UniqueAuthors4 0.3747 0.0339 11.05 <2e-16 ***
## UniqueAuthors5 0.4345 0.0326 13.31 <2e-16 ***
## Year1997 -0.0549 0.0759 -0.72 0.470
## Year1998 -0.0755 0.0761 -0.99 0.321
## Year1999 -0.0291 0.0773 -0.38 0.707
```

```

## Year2000      -0.1313      0.0729      -1.80      0.072 .
## Year2001      0.0243      0.0745      0.33      0.744
## Year2002      0.1417      0.0703      2.02      0.044 *
## Year2003     -0.1044      0.0672     -1.55      0.120
## Year2004      0.0446      0.0654      0.68      0.496
## Year2005      0.0370      0.0652      0.57      0.570
## Year2006     -0.0394      0.0635     -0.62      0.535
## Year2007     -0.1142      0.0670     -1.71      0.088 .
## Year2008      0.0583      0.0615      0.95      0.344
## Year2009      0.0705      0.0609      1.16      0.247
## Year2010      0.0184      0.0596      0.31      0.757
## Year2011      0.1005      0.0612      1.64      0.101
## Year2012      0.0381      0.0606      0.63      0.529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.0746, Adjusted R-squared:  0.0699
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 356 weights are ~= 1. The remaining 4005 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0606 0.8650 0.9500 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.29e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1      1.014
## LastAuthorFemale  1.011 1      1.006
## Year              1.034 16      1.001

```

## Residuals from first and last author



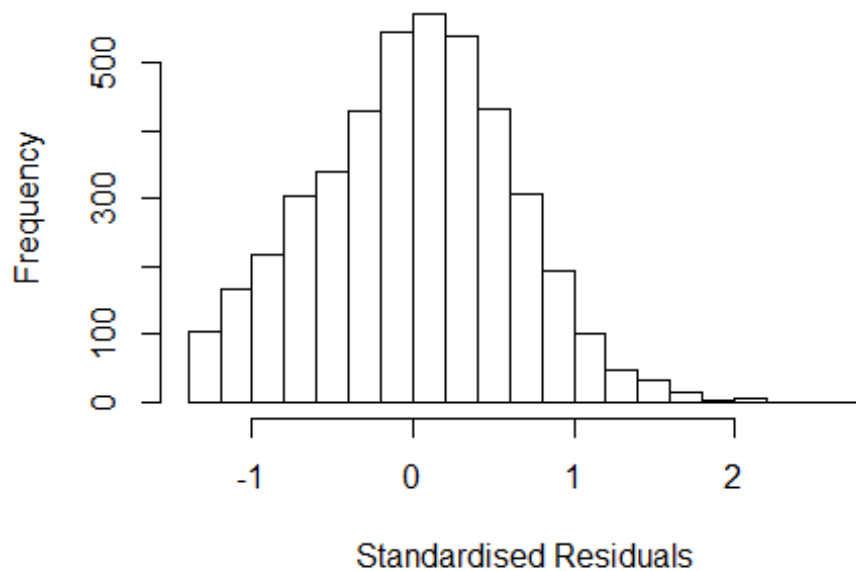
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2162 0031069035 3.593 1997    2210      3    2.599
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.264 -0.425  0.028  0.420  2.599
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.048193   0.055365  18.93  <2e-16 ***
## FirstAuthorFemale1 -0.005645   0.024380  -0.23   0.8169
## LastAuthorFemale1 -0.030850   0.035689  -0.86   0.3874
## Year1997        -0.054168   0.080903  -0.67   0.5032
## Year1998        -0.047327   0.081434  -0.58   0.5612
## Year1999        -0.023279   0.083233  -0.28   0.7797
## Year2000        -0.068987   0.077492  -0.89   0.3734
## Year2001         0.106078   0.076736   1.38   0.1669
## Year2002         0.215729   0.072648   2.97   0.0030 **
## Year2003        -0.005087   0.069531  -0.07   0.9417
## Year2004         0.155960   0.066752   2.34   0.0195 *
## Year2005         0.159278   0.067308   2.37   0.0180 *
```

```

## Year2006          0.073197    0.065336    1.12    0.2626
## Year2007          0.000836    0.069079    0.01    0.9903
## Year2008          0.165726    0.064728    2.56    0.0105 *
## Year2009          0.188204    0.063393    2.97    0.0030 **
## Year2010          0.141880    0.062141    2.28    0.0225 *
## Year2011          0.208728    0.064265    3.25    0.0012 **
## Year2012          0.180840    0.062768    2.88    0.0040 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.621
## Multiple R-squared:  0.0226, Adjusted R-squared:  0.0185
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 361 weights are ~= 1. The remaining 4000 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0407 0.8690 0.9490 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.29e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year              1.024 16      1.001

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2162 0031069035 3.593 1997    2210      3    2.599
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2612 -0.4241  0.0279  0.4213  2.6002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.04582    0.05525   18.93  <2e-16 ***
## FirstAuthorFemale1 -0.00930    0.02453   -0.38   0.7045
## Year1997       -0.05306    0.08089   -0.66   0.5119
## Year1998       -0.04664    0.08134   -0.57   0.5664
## Year1999       -0.02301    0.08322   -0.28   0.7822
## Year2000       -0.06877    0.07751   -0.89   0.3751
## Year2001        0.10657    0.07678    1.39   0.1652
## Year2002        0.21536    0.07270    2.96   0.0031 **
## Year2003       -0.00512    0.06957   -0.07   0.9413
## Year2004        0.15688    0.06676    2.35   0.0188 *
## Year2005        0.15911    0.06729    2.36   0.0181 *
## Year2006        0.07273    0.06536    1.11   0.2659
```

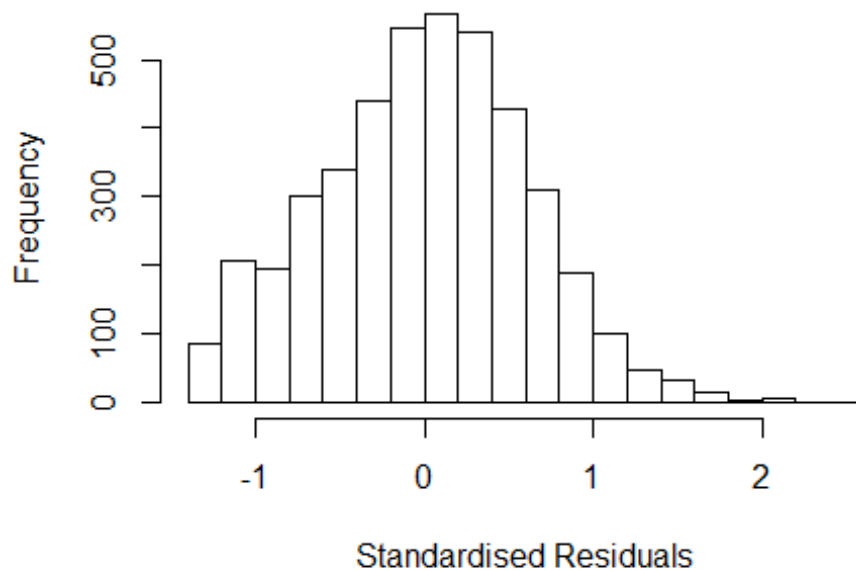
```

## Year2007          0.00112      0.06906      0.02      0.9870
## Year2008          0.16529      0.06477      2.55      0.0107 *
## Year2009          0.18845      0.06340      2.97      0.0030 **
## Year2010          0.14108      0.06219      2.27      0.0233 *
## Year2011          0.20826      0.06431      3.24      0.0012 **
## Year2012          0.18096      0.06281      2.88      0.0040 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.621
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.0186
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 353 weights are ~= 1. The remaining 4008 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0409 0.8700 0.9500 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.29e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.008 1      1.004
## Year      1.008 16      1.000

```



## Residuals from last author



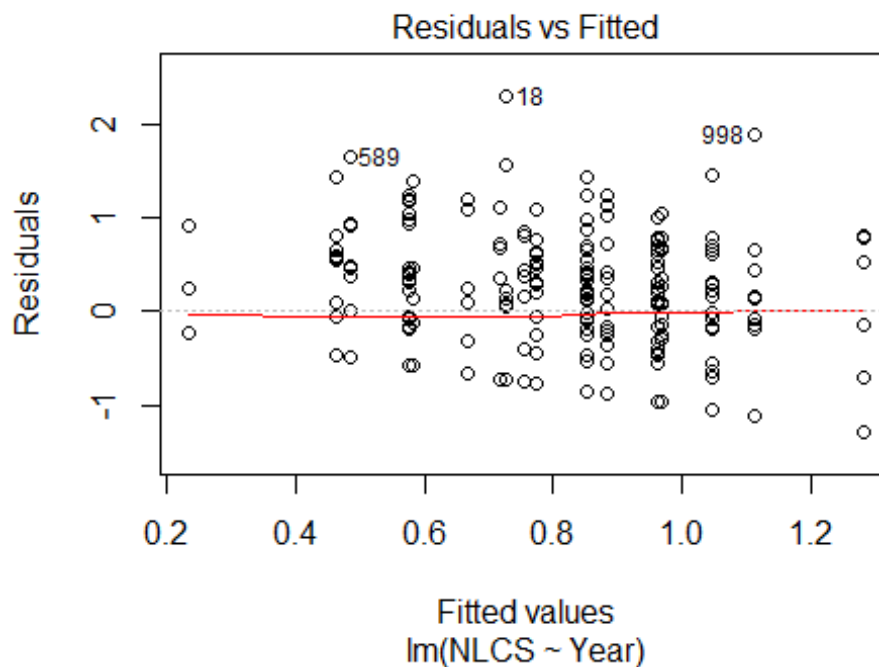
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2162 0031069035 3.593 1997      2210      3      2.599
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2631 -0.4254  0.0275  0.4208  2.5994
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.047785   0.055327   18.94  <2e-16 ***
## LastAuthorFemale1 -0.032004   0.035765   -0.89   0.3709
## Year1997        -0.054203   0.080892   -0.67   0.5029
## Year1998        -0.047517   0.081428   -0.58   0.5596
## Year1999        -0.023503   0.083220   -0.28   0.7776
## Year2000        -0.069270   0.077448   -0.89   0.3712
## Year2001         0.105676   0.076680    1.38   0.1682
## Year2002         0.215348   0.072607    2.97   0.0030 **
## Year2003        -0.005453   0.069491   -0.08   0.9375
## Year2004         0.155408   0.066629    2.33   0.0197 *
## Year2005         0.158714   0.067237    2.36   0.0183 *
## Year2006         0.072432   0.065228    1.11   0.2669
```

```

## Year2007          -0.000379    0.068732    -0.01    0.9956
## Year2008          0.165006    0.064641     2.55    0.0107 *
## Year2009          0.187478    0.063274     2.96    0.0031 **
## Year2010          0.141144    0.061996     2.28    0.0229 *
## Year2011          0.208177    0.064183     3.24    0.0012 **
## Year2012          0.180033    0.062607     2.88    0.0041 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.621
## Multiple R-squared:  0.0226, Adjusted R-squared:  0.0188
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 358 weights are ~= 1. The remaining 4003 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0406 0.8690 0.9490 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.29e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4361"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2501"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   74   70   83   83  101   70   72   82   79   58   39   58   34   19   16
## 2011 2012
##   10   12
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12    8    7    8   18   18   17   28   40   24   18   32   19   12    6
## 2011 2012

```

```
##      9      9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      8      6      7      6     15     15     15     28     37     19     13     28     17     10     6
## 2011 2012
##      9      8
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 9.8, df = 16, p-value = 0.9
```



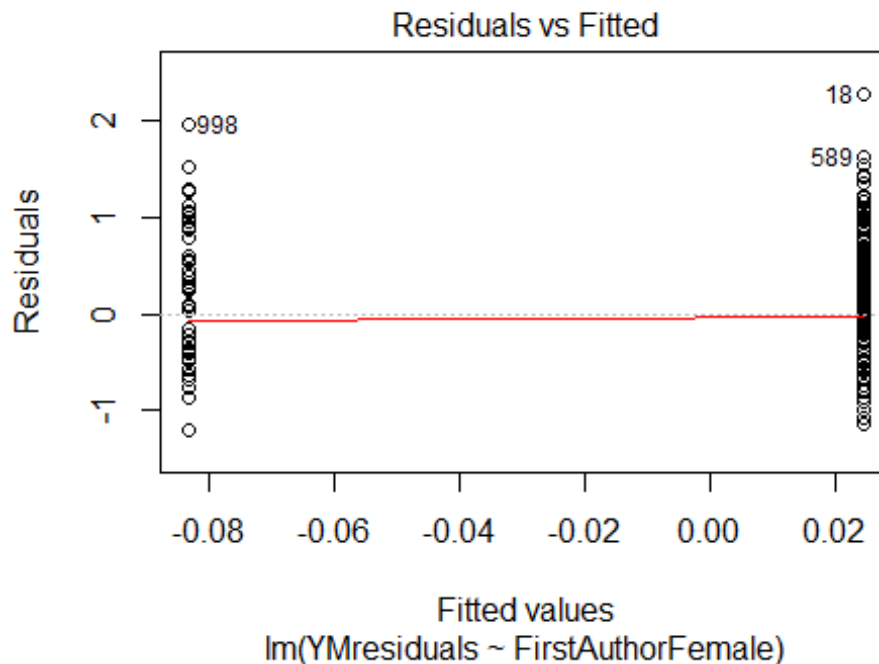
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.56, df = 1, p-value = 0.5

## [1] "Female first author team size 2018 geometric mean: 1.7370729389807"
## [1] "Male first author team size 2018 geometric mean: 2.71200674365364"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

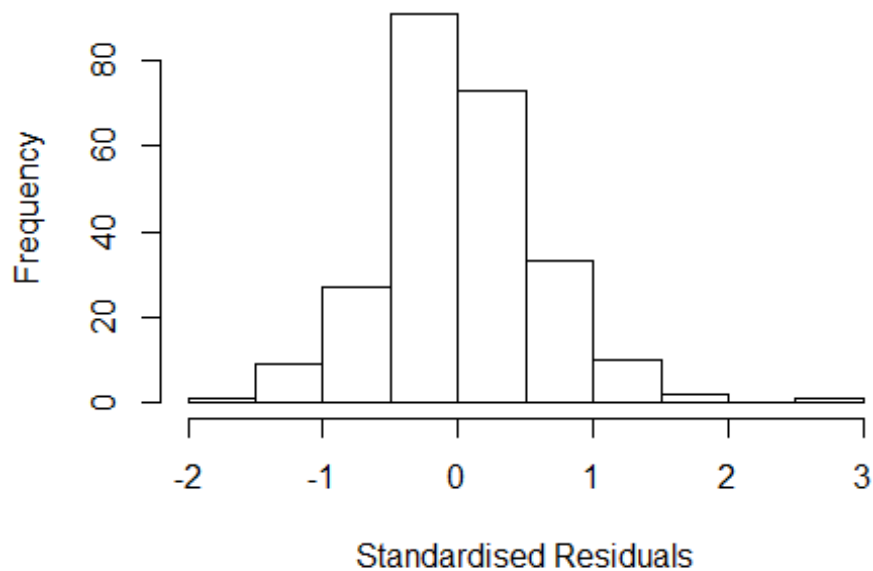
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 51, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.69172638514936"
## [1] "Male last author team size 2018 geometric mean: 2.83506334455035"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.054 1      2.460
## LastAuthorFemale 11.089 1      3.330
## UniqueAuthors    26.551 4      1.507
## Year              47.897 16     1.129
```

## Residuals from first and last author and team size



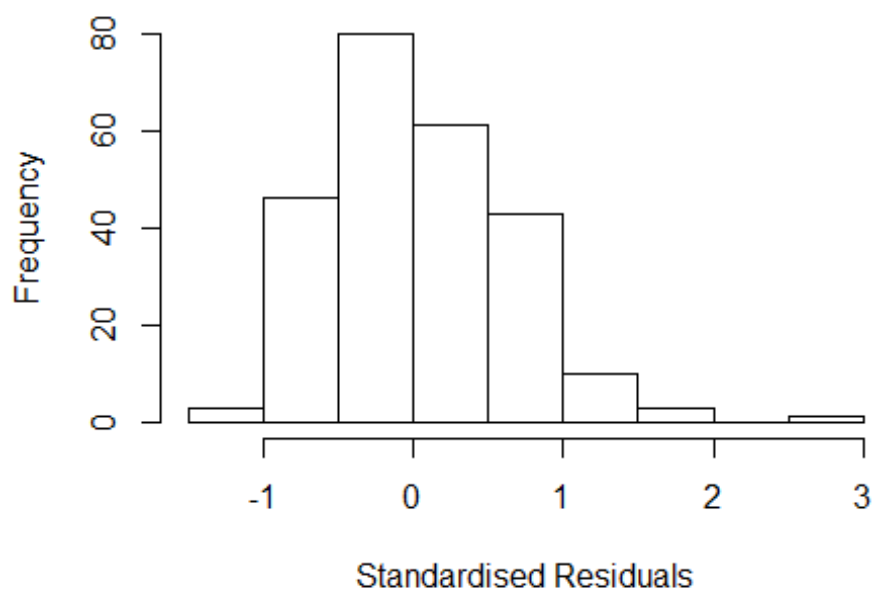
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 18 0030285260 3.028 1996    2501      1    2.633
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6935 -0.2798 -0.0713  0.3490  2.6329
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3951    0.1820   2.17 0.03102 *
## FirstAuthorFemale1 -0.0364    0.1428  -0.25 0.79905
## LastAuthorFemale1 -0.2416    0.1822  -1.33 0.18608
## UniqueAuthors2    0.7269    0.1965   3.70 0.00027 ***
## UniqueAuthors3    0.8551    0.1961   4.36 2.0e-05 ***
## UniqueAuthors4    0.5135    0.1895   2.71 0.00727 **
## UniqueAuthors5    0.6215    0.1532   4.06 6.9e-05 ***
## Year1997        -0.2327    0.3520  -0.66 0.50922
## Year1998        -0.2244    0.2240  -1.00 0.31736
## Year1999        -0.1153    0.3453  -0.33 0.73882
```

```

## Year2000          -0.0586      0.2645    -0.22  0.82494
## Year2001           0.0417      0.3112     0.13  0.89340
## Year2002          -0.2104      0.1898    -1.11  0.26873
## Year2003           0.1845      0.2193     0.84  0.40122
## Year2004           0.1152      0.2295     0.50  0.61634
## Year2005           0.1662      0.2477     0.67  0.50296
## Year2006          -0.2409      0.1945    -1.24  0.21661
## Year2007          -0.1558      0.1945    -0.80  0.42405
## Year2008           0.1200      0.2458     0.49  0.62595
## Year2009           0.5715      0.5206     1.10  0.27345
## Year2010           0.7157      0.4296     1.67  0.09715 .
## Year2011          -0.0278      0.2523    -0.11  0.91237
## Year2012           0.0921      0.2874     0.32  0.74878
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.42
## Multiple R-squared:  0.485, Adjusted R-squared:  0.434
## Convergence in 42 IRWLS iterations
##
## Robustness weights:
## observation 2 is an outlier with |weight| = 0 ( < 0.0004);
## 20 weights are ~= 1. The remaining 226 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0423 0.8240 0.9420 0.8520 0.9830 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           4.05e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.153 1 1.776
## LastAuthorFemale 2.950 1 1.718
## Year 1.986 16 1.022

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 18 0030285260 3.028 1996    2501      1    2.606
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1657 -0.4222 -0.0403  0.4543  2.6058
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4222    0.1819   2.32  0.0212 *
## FirstAuthorFemale1 -0.0121    0.1451  -0.08  0.9335
## LastAuthorFemale1 -0.3677    0.1480  -2.48  0.0137 *
## Year1997        -0.0142    0.2769  -0.05  0.9590
## Year1998        -0.2183    0.2313  -0.94  0.3461
## Year1999        -0.0240    0.4101  -0.06  0.9534
## Year2000         0.4830    0.2704   1.79  0.0754 .
## Year2001         0.5561    0.2522   2.20  0.0285 *
## Year2002         0.0294    0.2196   0.13  0.8935
## Year2003         0.6473    0.2247   2.88  0.0043 **
## Year2004         0.4467    0.2155   2.07  0.0393 *
## Year2005         0.7435    0.2463   3.02  0.0028 **
```

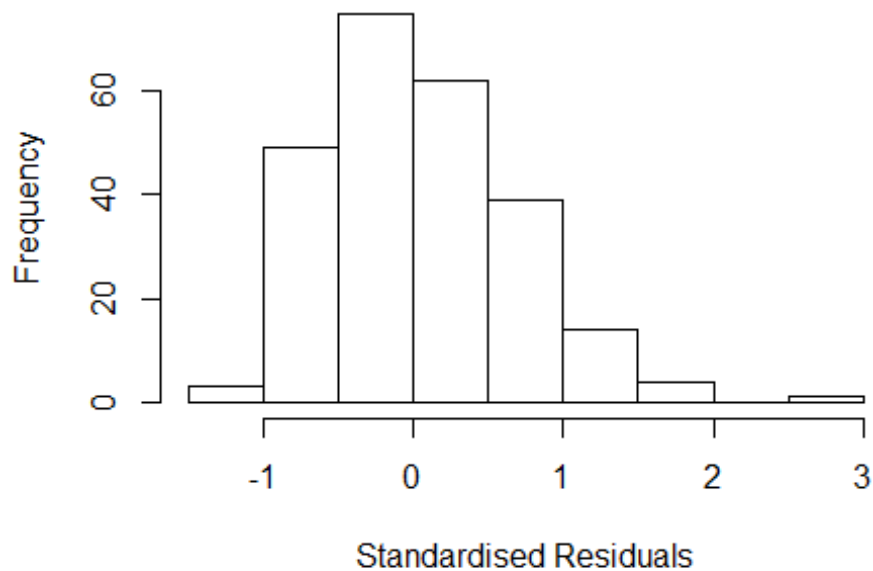
```

## Year2006          -0.0529      0.2293    -0.23    0.8178
## Year2007          0.1301      0.2057     0.63    0.5278
## Year2008          0.4380      0.2651     1.65    0.0999 .
## Year2009          0.5171      0.2765     1.87    0.0627 .
## Year2010          1.0148      0.3593     2.82    0.0052 **
## Year2011          0.3605      0.2833     1.27    0.2044
## Year2012          0.5388      0.3067     1.76    0.0802 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.218, Adjusted R-squared:  0.156
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0143 0.8540 0.9430 0.8990 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.05e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.485 1      1.219
## Year              1.485 16      1.012

```



## Residuals from first author



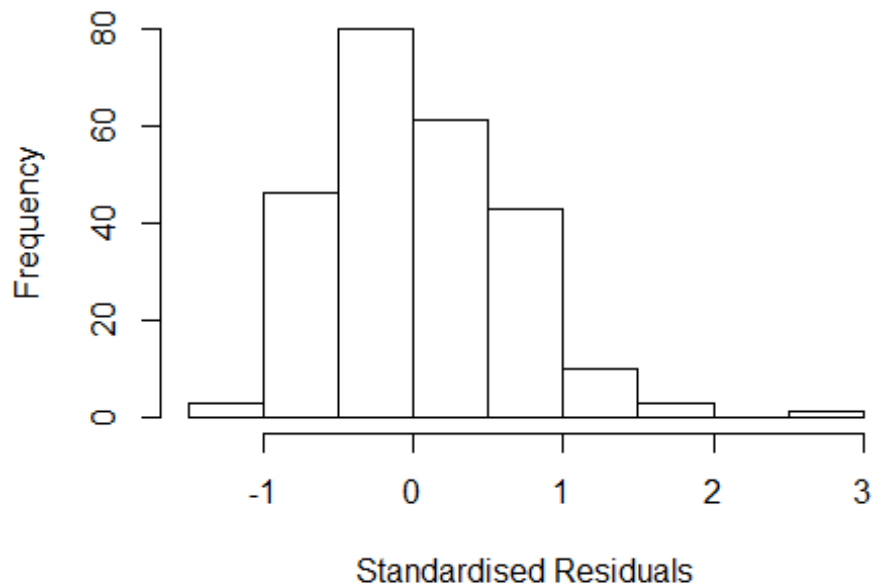
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 18 0030285260 3.028 1996    2501      1      2.606
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2034 -0.4045 -0.0566  0.4473  2.6286
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.39940    0.18520   2.16  0.0321 *
## FirstAuthorFemale1 -0.22793    0.11679  -1.95  0.0522 .
## Year1997      -0.06871    0.28102  -0.24  0.8071
## Year1998      -0.19448    0.23414  -0.83  0.4070
## Year1999       0.00506    0.41029   0.01  0.9902
## Year2000       0.48536    0.26574   1.83  0.0691 .
## Year2001       0.56918    0.25078   2.27  0.0242 *
## Year2002       0.01336    0.22862   0.06  0.9534
## Year2003       0.63569    0.22574   2.82  0.0053 **
## Year2004       0.42692    0.21813   1.96  0.0515 .
## Year2005       0.77907    0.24520   3.18  0.0017 **
## Year2006      -0.05181    0.23068  -0.22  0.8225
```

```

## Year2007          0.10358    0.21090    0.49    0.6238
## Year2008          0.45994    0.26637    1.73    0.0856 .
## Year2009          0.50913    0.29737    1.71    0.0882 .
## Year2010          1.03188    0.49160    2.10    0.0369 *
## Year2011          0.27579    0.28440    0.97    0.3332
## Year2012          0.47332    0.30958    1.53    0.1277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.602
## Multiple R-squared:  0.193, Adjusted R-squared:  0.133
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0176 0.8430 0.9500 0.9010 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.05e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.43 1      1.196
## Year      1.43 16      1.011

```

## Residuals from last author



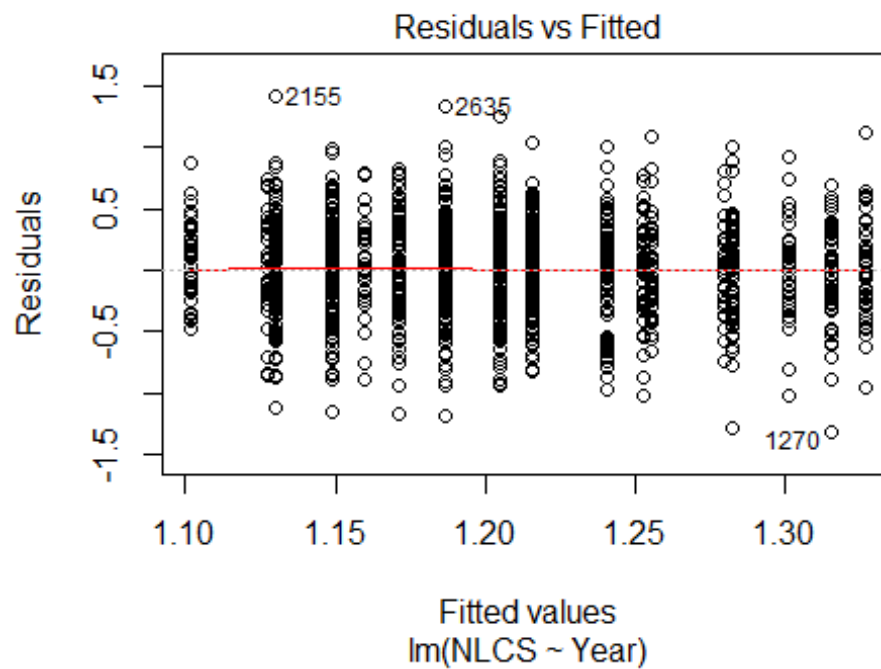
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 18 0030285260 3.028 1996    2501      1    2.606
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1634 -0.4197 -0.0339  0.4546  2.6083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4197    0.1794   2.34  0.02018 *
## LastAuthorFemale1 -0.3750    0.1056  -3.55  0.00046 ***
## Year1997      -0.0128    0.2715  -0.05  0.96233
## Year1998      -0.2164    0.2295  -0.94  0.34666
## Year1999      -0.0252    0.4103  -0.06  0.95103
## Year2000       0.4837    0.2693   1.80  0.07384 .
## Year2001       0.5586    0.2504   2.23  0.02665 *
## Year2002       0.0282    0.2169   0.13  0.89679
## Year2003       0.6480    0.2221   2.92  0.00388 **
## Year2004       0.4492    0.2123   2.12  0.03546 *
## Year2005       0.7437    0.2432   3.06  0.00249 **
## Year2006      -0.0545    0.2268  -0.24  0.81047
```

```

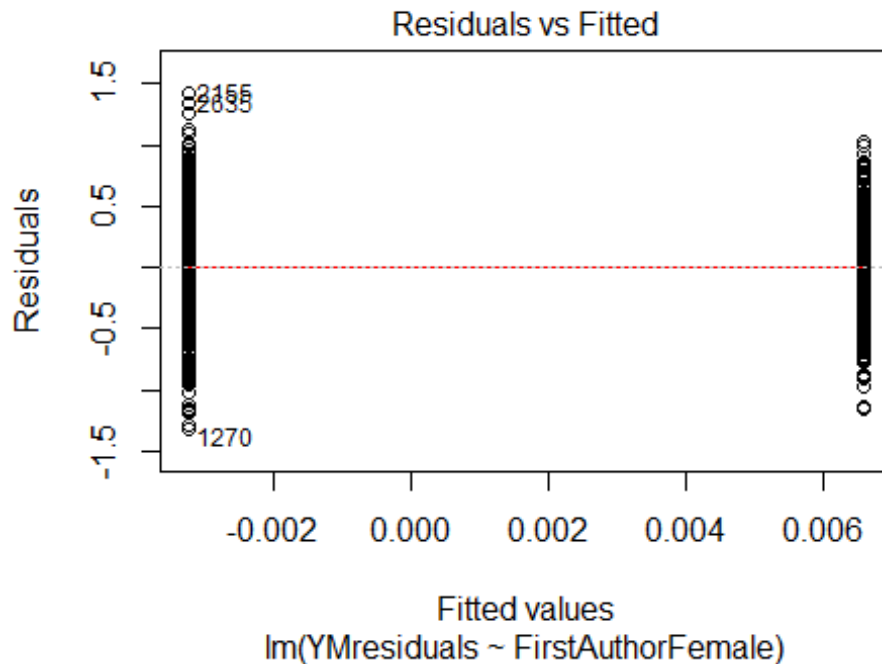
## Year2007          0.1293      0.2042      0.63  0.52727
## Year2008          0.4388      0.2616      1.68  0.09484 .
## Year2009          0.5209      0.2742      1.90  0.05869 .
## Year2010          1.0153      0.3573      2.84  0.00490 **
## Year2011          0.3604      0.2839      1.27  0.20552
## Year2012          0.5414      0.3066      1.77  0.07876 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.219, Adjusted R-squared:  0.161
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 227 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0102 0.8520 0.9390 0.8970 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.05e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 247"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2502"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 121 110 104 119 123 144 123 128 129 195 166 154 215 249 299
## 2011 2012
## 302 305
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 40 35 29 43 43 44 48 48 65 73 88 86 121 130 164
## 2011 2012

```

```
## 190 178
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 36 32 24 38 41 38 43 38 60 62 76 75 110 111 144
## 2011 2012
## 168 153
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 9.1, df = 16, p-value = 0.9
```

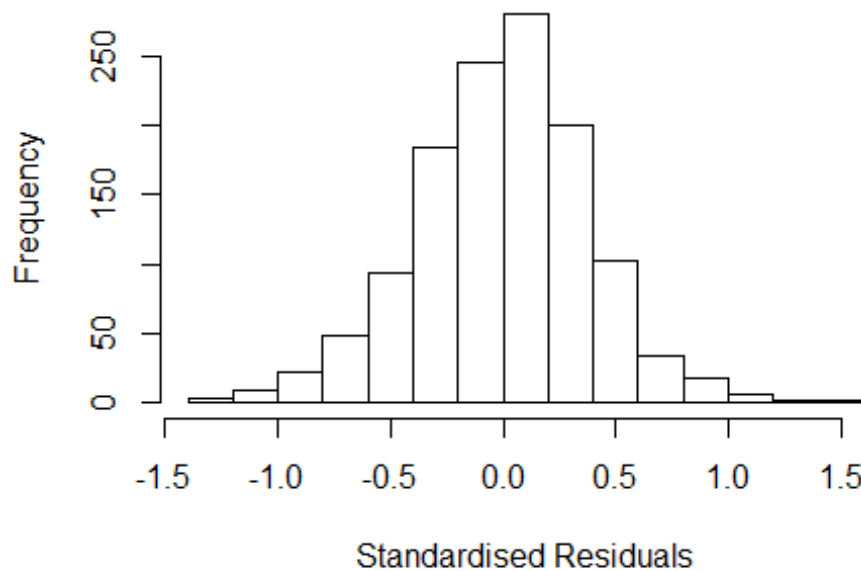


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 4.78544043027133"
## [1] "Male first author team size 2018 geometric mean: 3.57567265008653"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.33353120008107"
## [1] "Male last author team size 2018 geometric mean: 3.88233588571906"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.115 1 1.056
## LastAuthorFemale 1.101 1 1.049
## UniqueAuthors 1.456 4 1.048
## Year 1.484 16 1.012
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34628 -0.24442 0.00914 0.24389 1.58535
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98177 0.08004 12.27 < 2e-16 ***
## FirstAuthorFemale1 -0.01535 0.02317 -0.66 0.5079
## LastAuthorFemale1 0.00434 0.02960 0.15 0.8835
## UniqueAuthors2 0.16041 0.05737 2.80 0.0053 **
## UniqueAuthors3 0.16988 0.05551 3.06 0.0023 **
## UniqueAuthors4 0.16771 0.05713 2.94 0.0034 **
## UniqueAuthors5 0.27161 0.05502 4.94 9.1e-07 ***
## Year1997 -0.00258 0.09585 -0.03 0.9785
## Year1998 0.20238 0.09872 2.05 0.0406 *
## Year1999 -0.05162 0.09200 -0.56 0.5749
```

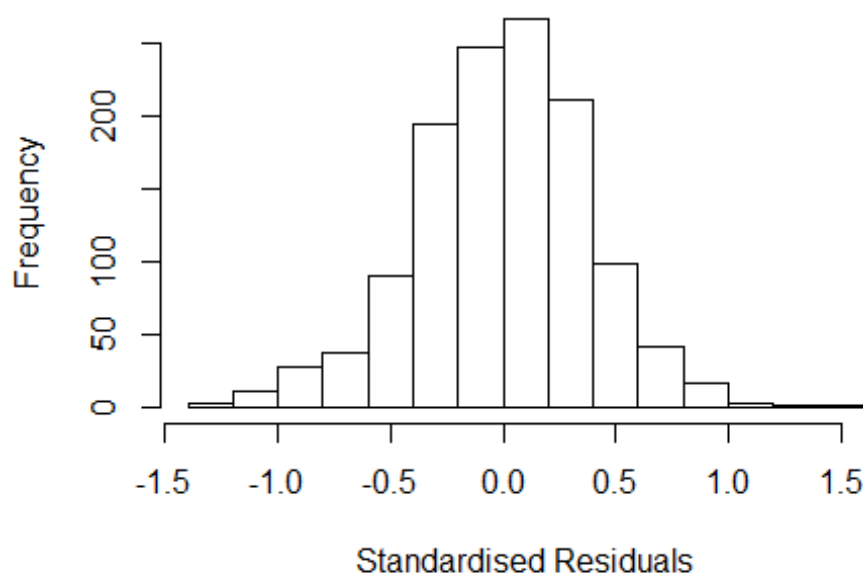
```

## Year2000      0.16975    0.08901    1.91    0.0568 .
## Year2001      0.18975    0.09264    2.05    0.0407 *
## Year2002      0.15689    0.09038    1.74    0.0828 .
## Year2003      0.04026    0.08778    0.46    0.6466
## Year2004      0.19975    0.08282    2.41    0.0160 *
## Year2005      0.12311    0.08626    1.43    0.1538
## Year2006      0.02974    0.08418    0.35    0.7240
## Year2007      0.09879    0.08232    1.20    0.2303
## Year2008     -0.02013    0.08057   -0.25    0.8028
## Year2009      0.03793    0.07849    0.48    0.6290
## Year2010      0.00296    0.07623    0.04    0.9690
## Year2011     -0.00506    0.07674   -0.07    0.9474
## Year2012      0.03256    0.07720    0.42    0.6733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.0704, Adjusted R-squared:  0.0537
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1142 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0125 0.8700 0.9470 0.8980 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1      1.029
## LastAuthorFemale  1.069 1      1.034
## Year              1.108 16      1.003

```



## Residuals from first and last author



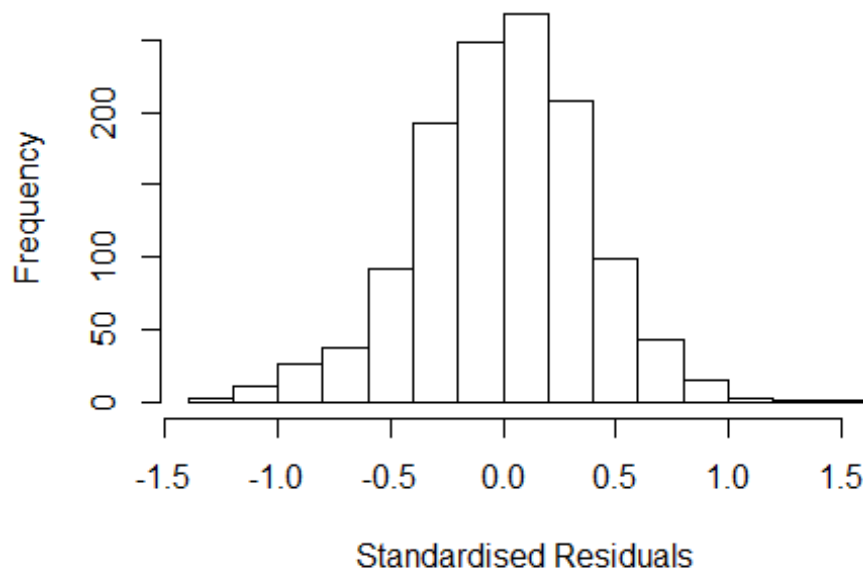
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3765 -0.2550  0.0115  0.2440  1.4039
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12716    0.07052   15.98  <2e-16 ***
## FirstAuthorFemale1 0.00595    0.02305    0.26  0.7963
## LastAuthorFemale1 0.00739    0.03032    0.24  0.8076
## Year1997         0.01611    0.09536    0.17  0.8659
## Year1998         0.20546    0.09989    2.06  0.0399 *
## Year1999        -0.04593    0.09117   -0.50  0.6145
## Year2000         0.17380    0.09025    1.93  0.0543 .
## Year2001         0.22874    0.09792    2.34  0.0197 *
## Year2002         0.16870    0.09232    1.83  0.0679 .
## Year2003         0.07318    0.08738    0.84  0.4025
## Year2004         0.24197    0.08429    2.87  0.0042 **
## Year2005         0.17142    0.08546    2.01  0.0451 *
```

```

## Year2006          0.07734      0.08550      0.90      0.3659
## Year2007          0.15187      0.08387      1.81      0.0704 .
## Year2008          0.01590      0.08076      0.20      0.8440
## Year2009          0.08443      0.07936      1.06      0.2876
## Year2010          0.05166      0.07716      0.67      0.5033
## Year2011          0.04385      0.07745      0.57      0.5714
## Year2012          0.09096      0.07798      1.17      0.2436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0336, Adjusted R-squared:  0.0195
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 1154 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.120  0.877   0.951   0.901   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.049 1          1.024
## Year              1.049 16          1.002

```

## Residuals from first author



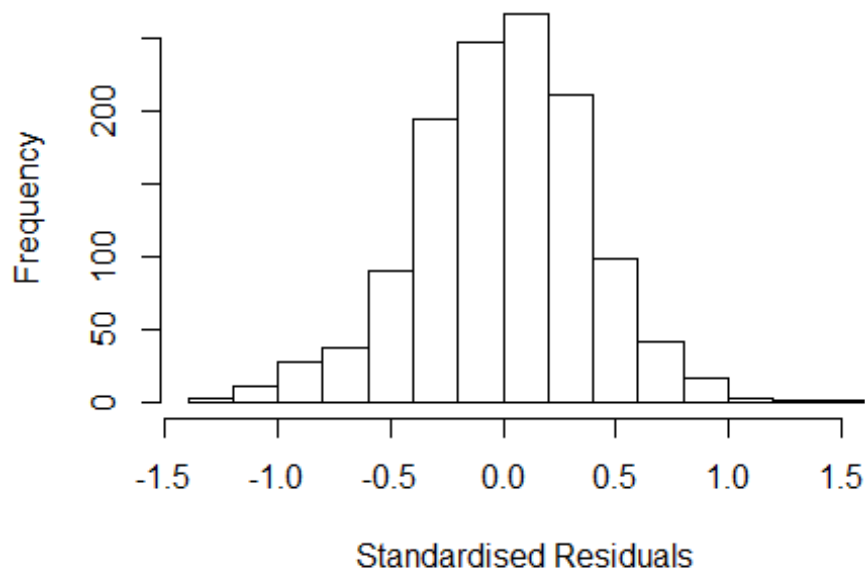
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3693 -0.2545  0.0104  0.2442  1.4032
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12776    0.07047   16.00  <2e-16 ***
## FirstAuthorFemale1 0.00664    0.02296    0.29  0.7723
## Year1997        0.01606    0.09525    0.17  0.8661
## Year1998        0.20517    0.09986    2.05  0.0401 *
## Year1999       -0.04565    0.09110   -0.50  0.6164
## Year2000        0.17421    0.09011    1.93  0.0534 .
## Year2001        0.22975    0.09764    2.35  0.0188 *
## Year2002        0.16937    0.09236    1.83  0.0669 .
## Year2003        0.07320    0.08727    0.84  0.4017
## Year2004        0.24156    0.08425    2.87  0.0042 **
## Year2005        0.17173    0.08541    2.01  0.0446 *
## Year2006        0.07778    0.08546    0.91  0.3629
```

```

## Year2007          0.15271    0.08363    1.83    0.0681 .
## Year2008          0.01603    0.08073    0.20    0.8426
## Year2009          0.08493    0.07926    1.07    0.2842
## Year2010          0.05186    0.07711    0.67    0.5014
## Year2011          0.04480    0.07718    0.58    0.5617
## Year2012          0.09168    0.07776    1.18    0.2386
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0335, Adjusted R-squared:  0.0202
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 1151 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.119  0.878  0.951  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.059 1      1.029
## Year              1.059 16      1.002

```

## Residuals from last author



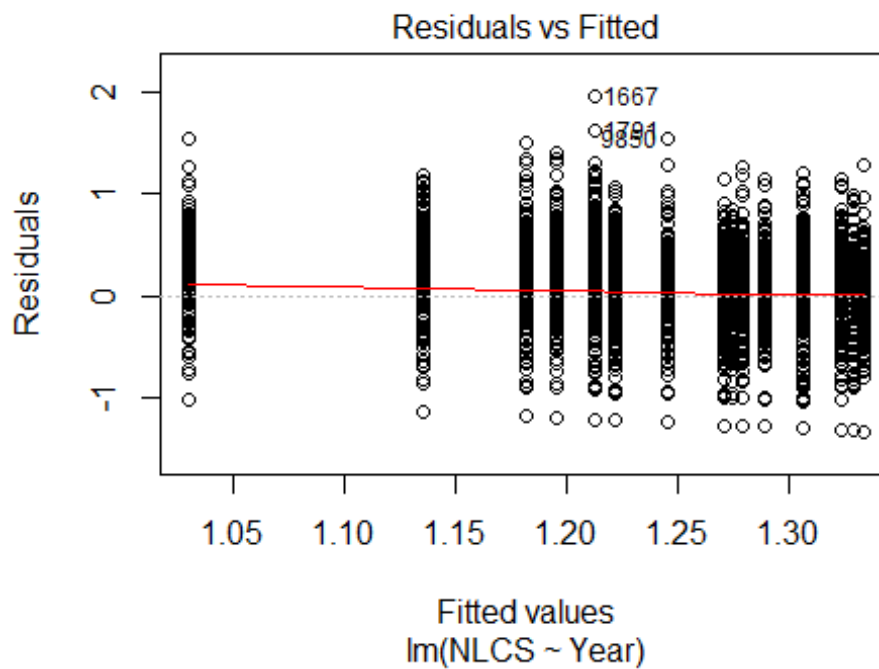
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3796 -0.2571  0.0126  0.2439  1.4021
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1281    0.0707   15.95  <2e-16 ***
## LastAuthorFemale1  0.0083    0.0302    0.27   0.784
## Year1997          0.0177    0.0951    0.19   0.852
## Year1998          0.2054    0.1000    2.05   0.040 *
## Year1999         -0.0459    0.0914   -0.50   0.616
## Year2000          0.1745    0.0903    1.93   0.054 .
## Year2001          0.2298    0.0977    2.35   0.019 *
## Year2002          0.1701    0.0920    1.85   0.065 .
## Year2003          0.0743    0.0875    0.85   0.396
## Year2004          0.2433    0.0843    2.89   0.004 **
## Year2005          0.1722    0.0855    2.01   0.044 *
## Year2006          0.0784    0.0854    0.92   0.359
```

```

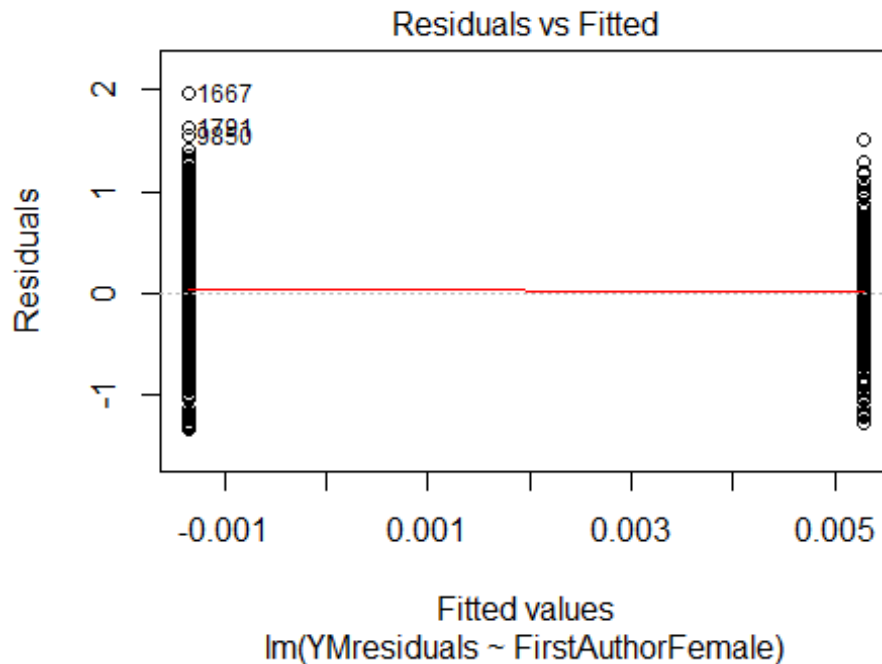
## Year2007          0.1531      0.0837      1.83      0.068 .
## Year2008          0.0168      0.0808      0.21      0.835
## Year2009          0.0858      0.0791      1.08      0.278
## Year2010          0.0529      0.0770      0.69      0.492
## Year2011          0.0448      0.0774      0.58      0.563
## Year2012          0.0918      0.0780      1.18      0.240
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0336, Adjusted R-squared:  0.0203
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1157 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.878  0.951  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1249"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2503"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 627 592 657 600 611 511 516 475 433 387 411 389 358 411 405
## 2011 2012
## 454 460
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 294 269 302 270 180 113 241 219 206 195 191 179 182 203 206
## 2011 2012

```

```
## 249 225
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 255 233 270 247 164 97 217 195 188 167 176 157 166 185 186
## 2011 2012
## 221 203
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 180, df = 16, p-value <2e-16
```



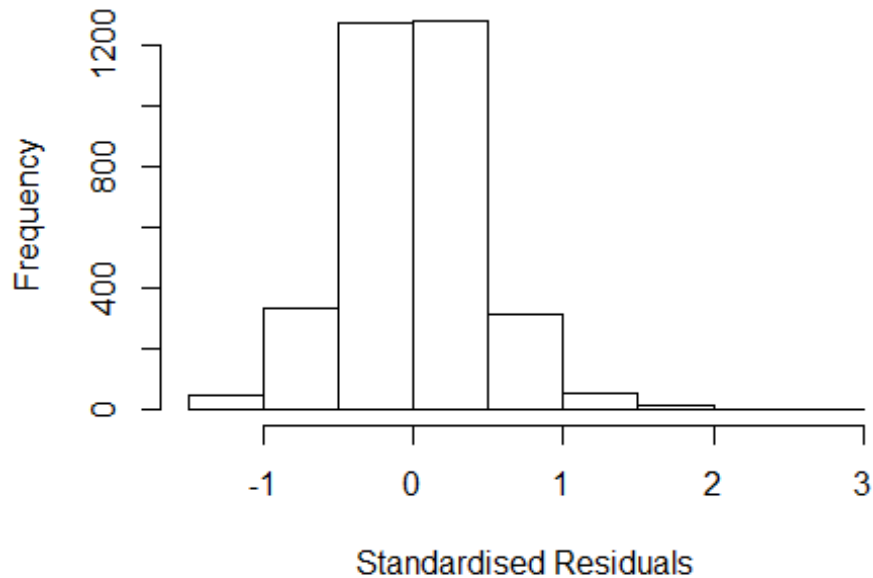
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 18, df = 1, p-value = 2e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.54743733782218"
## [1] "Male first author team size 2018 geometric mean: 3.68965807839708"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.22855395668087"
## [1] "Male last author team size 2018 geometric mean: 3.86296781778853"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1 1.023
## LastAuthorFemale 1.018 1 1.009
## UniqueAuthors 1.225 4 1.026
## Year 1.279 16 1.008
```



## Residuals from first and last author and team size



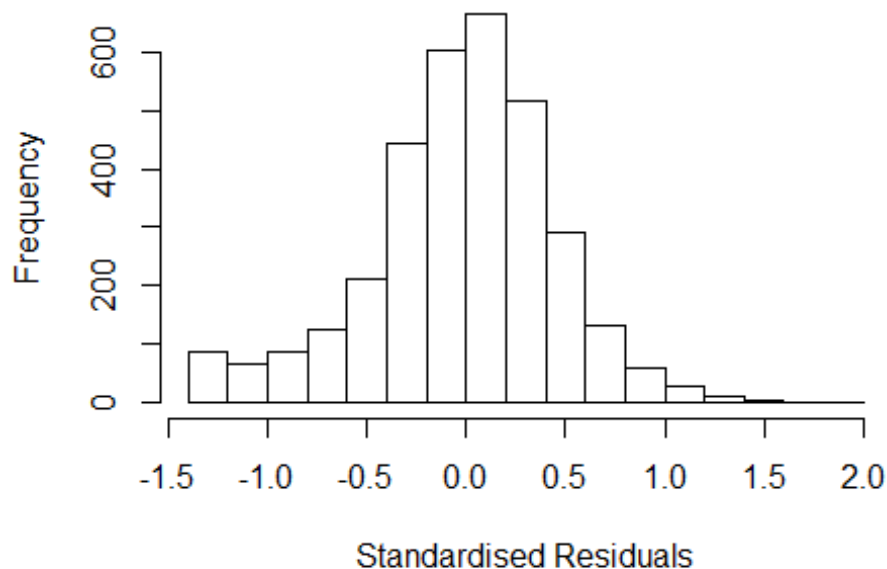
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1667 0032120696 3.174 1998      2503      2      2.604
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.351828 -0.261203  0.000238  0.267836  2.603578
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.539529   0.057961    9.31  <2e-16 ***
## FirstAuthorFemale1 -0.027446   0.017353   -1.58   0.1138
## LastAuthorFemale1  0.007113   0.023142    0.31   0.7586
## UniqueAuthors2     0.712072   0.057085   12.47  <2e-16 ***
## UniqueAuthors3     0.720932   0.054861   13.14  <2e-16 ***
## UniqueAuthors4     0.786908   0.054795   14.36  <2e-16 ***
## UniqueAuthors5     0.760378   0.054739   13.89  <2e-16 ***
## Year1997          -0.022079   0.044782   -0.49   0.6220
## Year1998           0.030892   0.041379    0.75   0.4554
## Year1999          -0.020264   0.042383   -0.48   0.6326
```

```

## Year2000      -0.093965    0.050765    -1.85    0.0643 .
## Year2001      0.131176    0.050112     2.62    0.0089 **
## Year2002      0.091367    0.042679     2.14    0.0324 *
## Year2003     -0.003380    0.039970    -0.08    0.9326
## Year2004      0.058587    0.042108     1.39    0.1642
## Year2005      0.055681    0.041395     1.35    0.1787
## Year2006      0.052324    0.041754     1.25    0.2102
## Year2007      0.020288    0.043301     0.47    0.6394
## Year2008     -0.000734    0.041949    -0.02    0.9860
## Year2009      0.046853    0.041869     1.12    0.2632
## Year2010      0.005094    0.042186     0.12    0.9039
## Year2011      0.021411    0.039052     0.55    0.5835
## Year2012     -0.020615    0.040191    -0.51    0.6080
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.22,   Adjusted R-squared:  0.214
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 7 observations c(233,237,535,638,1159,2460,3082)
## are outliers with |weight| = 0 ( < 3e-05);
## 283 weights are ~ = 1. The remaining 3037 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0022 0.8520 0.9480 0.8860 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.01e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1 1.023
## LastAuthorFemale 1.016 1 1.008
## Year 1.053 16 1.002

```

## Residuals from first and last author



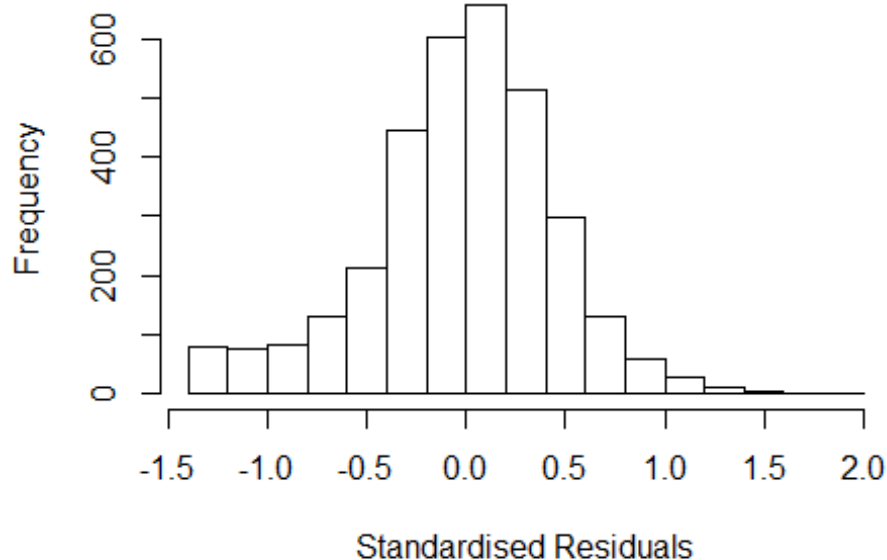
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3548 -0.2746 0.0125 0.2652 1.9303
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18833 0.03089 38.47 < 2e-16 ***
## FirstAuthorFemale1 0.00117 0.01809 0.06 0.94829
## LastAuthorFemale1 0.02722 0.02444 1.11 0.26554
## Year1997 0.03380 0.04841 0.70 0.48514
## Year1998 0.05542 0.04667 1.19 0.23514
## Year1999 0.00445 0.04832 0.09 0.92665
## Year2000 -0.09473 0.06797 -1.39 0.16351
## Year2001 0.16649 0.05191 3.21 0.00135 **
## Year2002 0.14836 0.04244 3.50 0.00048 ***
## Year2003 0.06550 0.04030 1.63 0.10413
## Year2004 0.12160 0.04358 2.79 0.00530 **
## Year2005 0.11912 0.04242 2.81 0.00501 **
```

```

## Year2006          0.13806      0.04184      3.30  0.00098 ***
## Year2007          0.09492      0.04527      2.10  0.03609 *
## Year2008          0.09068      0.04346      2.09  0.03702 *
## Year2009          0.14442      0.04325      3.34  0.00085 ***
## Year2010          0.08293      0.04279      1.94  0.05267 .
## Year2011          0.08375      0.03969      2.11  0.03492 *
## Year2012          0.04176      0.04115      1.01  0.31020
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0195, Adjusted R-squared:  0.0142
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 638 is an outlier with |weight| = 0 ( < 3e-05);
## 291 weights are ~ = 1. The remaining 3035 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0911 0.8550 0.9510 0.8800 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1          1.020
## Year              1.041 16          1.001

```

## Residuals from first author



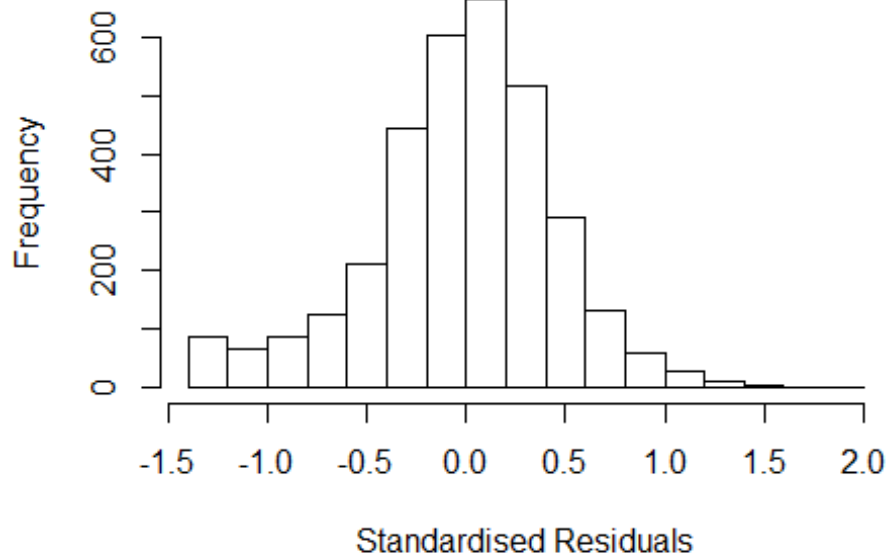
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3587 -0.2710 0.0139 0.2696 1.9274
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19068 0.03074 38.73 < 2e-16 ***
## FirstAuthorFemale1 0.00170 0.01814 0.09 0.92536
## Year1997 0.03329 0.04843 0.69 0.49195
## Year1998 0.05596 0.04670 1.20 0.23092
## Year1999 0.00513 0.04830 0.11 0.91544
## Year2000 -0.09589 0.06797 -1.41 0.15844
## Year2001 0.16797 0.05190 3.24 0.00122 **
## Year2002 0.14939 0.04236 3.53 0.00043 ***
## Year2003 0.06574 0.04027 1.63 0.10269
## Year2004 0.12106 0.04360 2.78 0.00552 **
## Year2005 0.11894 0.04240 2.80 0.00506 **
## Year2006 0.13866 0.04185 3.31 0.00093 ***
```

```

## Year2007          0.09474    0.04533    2.09  0.03669 *
## Year2008          0.09034    0.04344    2.08  0.03764 *
## Year2009          0.14489    0.04324    3.35  0.00081 ***
## Year2010          0.08397    0.04277    1.96  0.04967 *
## Year2011          0.08471    0.03972    2.13  0.03300 *
## Year2012          0.04261    0.04105    1.04  0.29930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0192, Adjusted R-squared:  0.0141
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 638 is an outlier with |weight| = 0 ( < 3e-05);
## 299 weights are ~= 1. The remaining 3027 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0932 0.8540 0.9520 0.8790 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.01e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.011 1          1.006
## Year          1.011 16          1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3550 -0.2748 0.0127 0.2650 1.9300
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18847 0.03082 38.57 < 2e-16 ***
## LastAuthorFemale1 0.02726 0.02449 1.11 0.26577
## Year1997 0.03392 0.04836 0.70 0.48306
## Year1998 0.05557 0.04658 1.19 0.23294
## Year1999 0.00453 0.04831 0.09 0.92538
## Year2000 -0.09462 0.06798 -1.39 0.16404
## Year2001 0.16651 0.05191 3.21 0.00135 **
## Year2002 0.14845 0.04238 3.50 0.00047 ***
## Year2003 0.06561 0.04027 1.63 0.10329
## Year2004 0.12171 0.04352 2.80 0.00519 **
## Year2005 0.11927 0.04238 2.81 0.00492 **
## Year2006 0.13819 0.04178 3.31 0.00095 ***
```

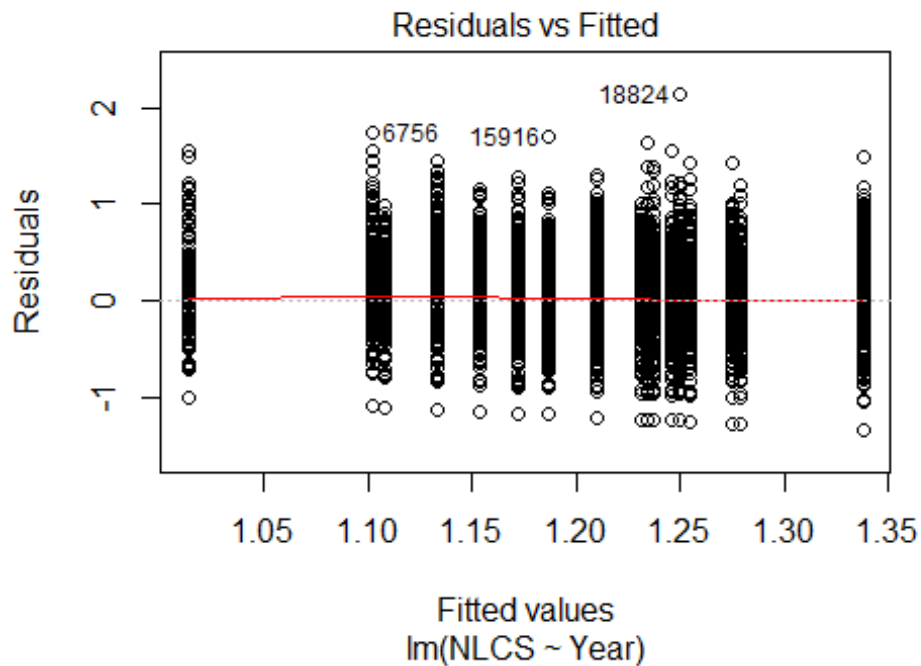
```

## Year2007      0.09513      0.04514      2.11  0.03513 *
## Year2008      0.09086      0.04330      2.10  0.03595 *
## Year2009      0.14462      0.04314      3.35  0.00081 ***
## Year2010      0.08308      0.04272      1.94  0.05189 .
## Year2011      0.08405      0.03930      2.14  0.03254 *
## Year2012      0.04198      0.04096      1.02  0.30544
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.0195, Adjusted R-squared:  0.0145
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 638 is an outlier with |weight| = 0 ( < 3e-05);
## 295 weights are ~= 1. The remaining 3031 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.091  0.855  0.951  0.879  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3327"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2504"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1003 717 1011 979 1216 1181 1006 759 655 663 740 783 850 881 918
## 2011 2012
## 887 865
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 162 141 338 240 269 278 427 298 277 268 289 328 361 373 414

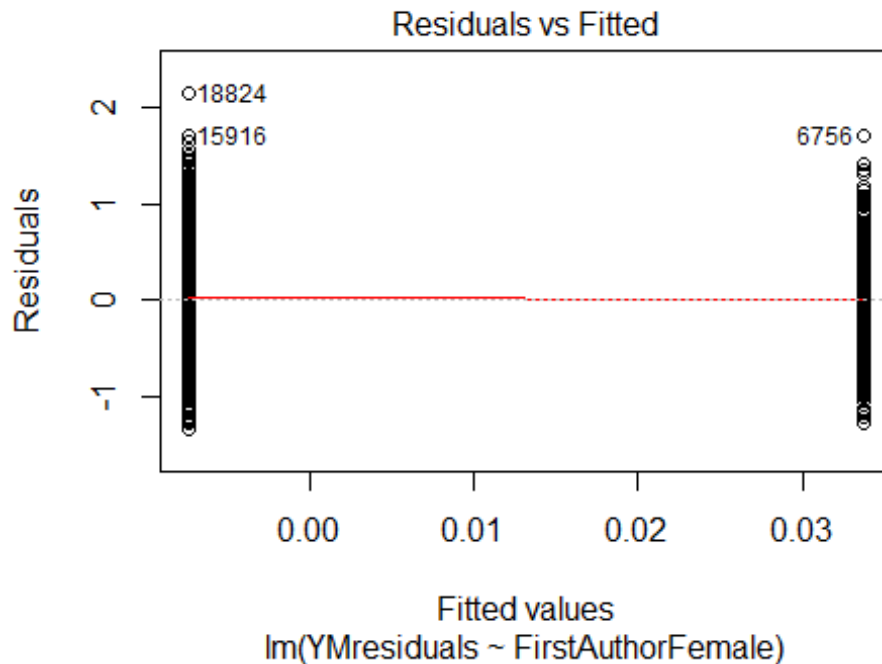
```



```
## 2011 2012
## 429 409
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 137 116 290 213 237 236 354 252 240 230 242 278 320 320 356
## 2011 2012
## 378 355
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 160, df = 16, p-value <2e-16
```

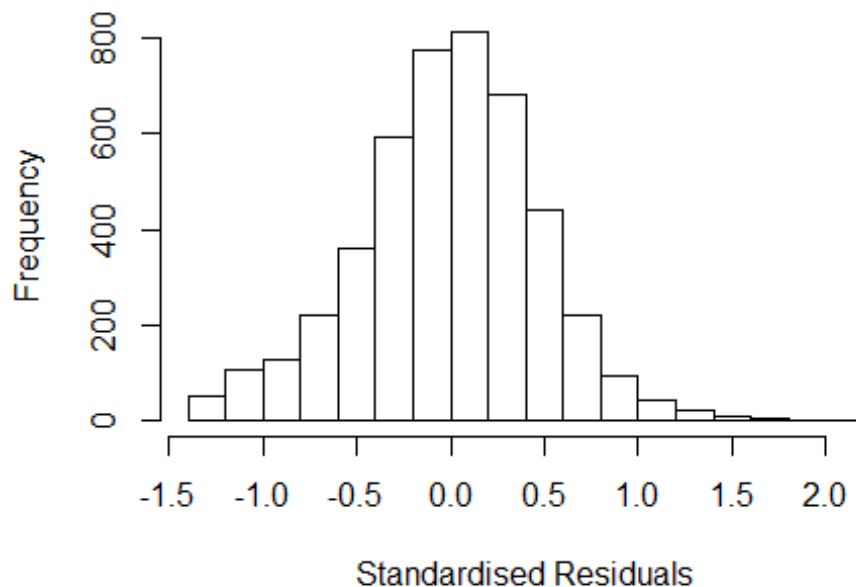


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 21, df = 1, p-value = 5e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.50040961650794"
## [1] "Male first author team size 2018 geometric mean: 3.71122704552741"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4800, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.79066987690197"
## [1] "Male last author team size 2018 geometric mean: 3.84881269431829"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1          1.030
## LastAuthorFemale  1.041 1          1.020
## UniqueAuthors    1.164 4          1.019
## Year             1.196 16          1.006
```

## Residuals from first and last author and team size



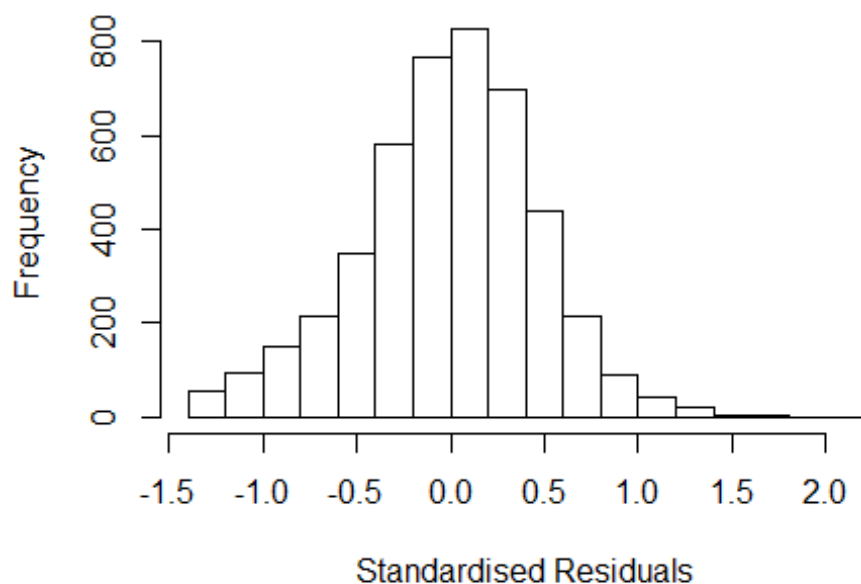
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.39709 -0.28986  0.00903  0.29573  2.14792
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.87338    0.06526   13.38 < 2e-16 ***
## FirstAuthorFemale1 0.02308    0.01674    1.38  0.16811
## LastAuthorFemale1 0.00145    0.02247    0.06  0.94839
## UniqueAuthors2    0.09883    0.03811    2.59  0.00953 **
## UniqueAuthors3    0.11456    0.03769    3.04  0.00238 **
## UniqueAuthors4    0.14560    0.03839    3.79  0.00015 ***
## UniqueAuthors5    0.17344    0.03689    4.70  2.7e-06 ***
## Year1997          0.17689    0.07751    2.28  0.02254 *
## Year1998          0.34881    0.06635    5.26  1.5e-07 ***
## Year1999          0.28446    0.06628    4.29  1.8e-05 ***
```

```

## Year2000          0.20223      0.06949      2.91  0.00363 **
## Year2001          0.13949      0.07641      1.83  0.06799 .
## Year2002          0.19869      0.06723      2.96  0.00314 **
## Year2003          0.25618      0.06485      3.95  7.9e-05 ***
## Year2004          0.25209      0.06539      3.86  0.00012 ***
## Year2005          0.27169      0.06610      4.11  4.0e-05 ***
## Year2006          0.24328      0.06607      3.68  0.00023 ***
## Year2007          0.24928      0.06453      3.86  0.00011 ***
## Year2008          0.24383      0.06312      3.86  0.00011 ***
## Year2009          0.21260      0.06547      3.25  0.00117 **
## Year2010          0.21011      0.06429      3.27  0.00109 **
## Year2011          0.25315      0.06358      3.98  7.0e-05 ***
## Year2012          0.18685      0.06421      2.91  0.00363 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.0298, Adjusted R-squared:  0.0251
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 4023 is an outlier with |weight| = 0 ( < 2.2e-05);
## 396 weights are ~= 1. The remaining 4157 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0911 0.8620 0.9500 0.8920 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.20e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev mts compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1          1.019
## LastAuthorFemale 1.029 1          1.014
## Year 1.068 16          1.002

```

## Residuals from first and last author



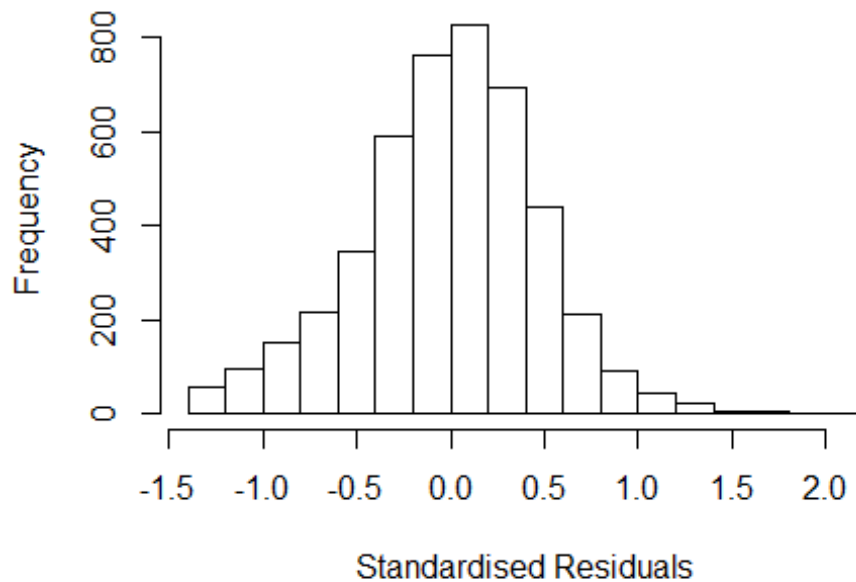
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3490 -0.2903 0.0135 0.2972 2.1382
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96669 0.05927 16.31 < 2e-16 ***
## FirstAuthorFemale1 0.03536 0.01666 2.12 0.03379 *
## LastAuthorFemale1 0.00822 0.02246 0.37 0.71451
## Year1997 0.18728 0.07734 2.42 0.01550 *
## Year1998 0.37412 0.06567 5.70 1.3e-08 ***
## Year1999 0.30992 0.06604 4.69 2.8e-06 ***
## Year2000 0.22223 0.06927 3.21 0.00135 **
## Year2001 0.16708 0.07577 2.21 0.02749 *
## Year2002 0.22370 0.06669 3.35 0.00080 ***
## Year2003 0.28893 0.06428 4.50 7.1e-06 ***
## Year2004 0.27758 0.06515 4.26 2.1e-05 ***
## Year2005 0.30312 0.06593 4.60 4.4e-06 ***
```

```

## Year2006          0.27449      0.06564      4.18  2.9e-05 ***
## Year2007          0.28056      0.06406      4.38  1.2e-05 ***
## Year2008          0.27175      0.06301      4.31  1.6e-05 ***
## Year2009          0.24748      0.06511      3.80  0.00015 ***
## Year2010          0.24064      0.06401      3.76  0.00017 ***
## Year2011          0.28414      0.06340      4.48  7.6e-06 ***
## Year2012          0.22179      0.06384      3.47  0.00052 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0157
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 4023 is an outlier with |weight| = 0 ( < 2.2e-05);
## 370 weights are ~ = 1. The remaining 4183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.109  0.859  0.951   0.892   0.985   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          2.20e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1           1.019
## Year              1.038 16           1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.341 -0.290  0.014  0.296  2.137
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9673    0.0592   16.33 < 2e-16 ***
## FirstAuthorFemale1 0.0358    0.0167    2.15  0.03191 *
## Year1997        0.1871    0.0774    2.42  0.01563 *
## Year1998        0.3742    0.0657    5.70  1.3e-08 ***
## Year1999        0.3103    0.0661    4.70  2.7e-06 ***
## Year2000        0.2225    0.0693    3.21  0.00134 **
## Year2001        0.1676    0.0757    2.21  0.02698 *
## Year2002        0.2242    0.0667    3.36  0.00078 ***
## Year2003        0.2891    0.0643    4.50  7.1e-06 ***
## Year2004        0.2777    0.0652    4.26  2.1e-05 ***
## Year2005        0.3032    0.0660    4.60  4.4e-06 ***
## Year2006        0.2746    0.0657    4.18  2.9e-05 ***
```

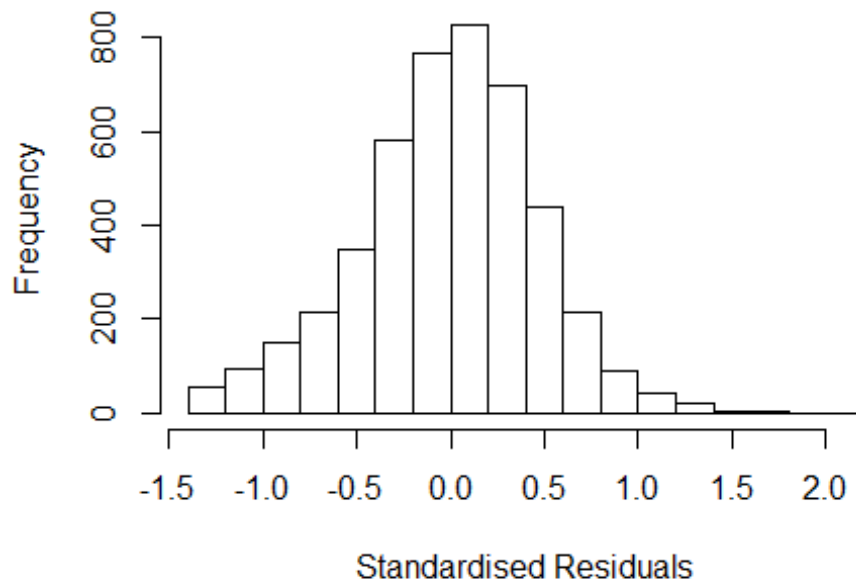
```

## Year2007          0.2807      0.0641      4.38  1.2e-05 ***
## Year2008          0.2716      0.0630      4.31  1.7e-05 ***
## Year2009          0.2475      0.0651      3.80  0.00015 ***
## Year2010          0.2408      0.0640      3.76  0.00017 ***
## Year2011          0.2844      0.0634      4.48  7.5e-06 ***
## Year2012          0.2222      0.0639      3.48  0.00051 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0196, Adjusted R-squared:  0.0159
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 4023 is an outlier with |weight| = 0 ( < 2.2e-05);
## 374 weights are ~= 1. The remaining 4179 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.109  0.859  0.951  0.892  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1      1.014
## Year      1.028 16      1.001

```



## Residuals from last author



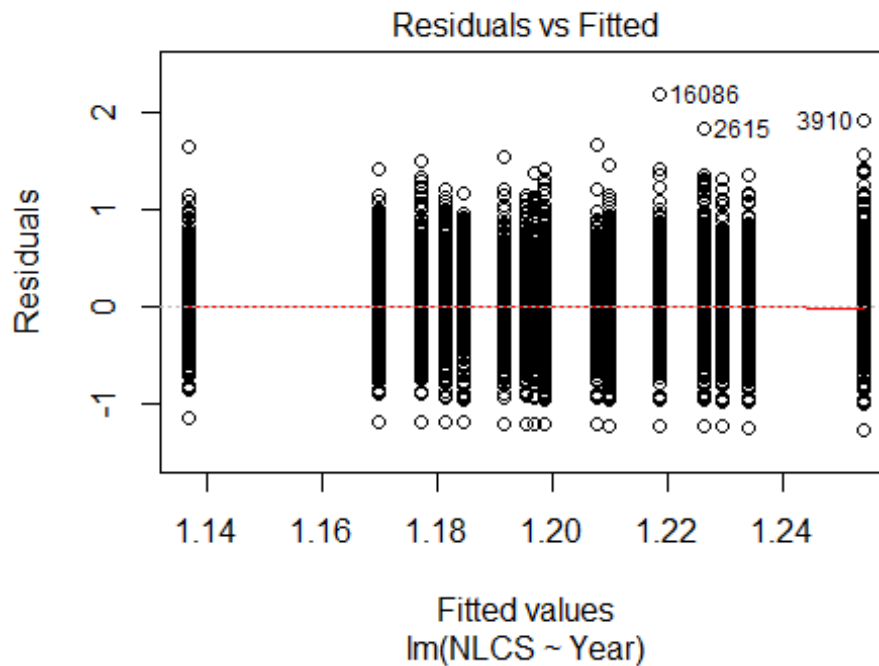
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3590 -0.2952 0.0159 0.2974 2.1279
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9711 0.0592 16.41 < 2e-16 ***
## LastAuthorFemale1 0.0118 0.0225 0.52 0.60017
## Year1997 0.1902 0.0773 2.46 0.01396 *
## Year1998 0.3761 0.0656 5.74 1.0e-08 ***
## Year1999 0.3088 0.0660 4.68 3.0e-06 ***
## Year2000 0.2214 0.0693 3.19 0.00141 **
## Year2001 0.1674 0.0756 2.21 0.02689 *
## Year2002 0.2245 0.0666 3.37 0.00076 ***
## Year2003 0.2905 0.0642 4.52 6.3e-06 ***
## Year2004 0.2781 0.0651 4.27 2.0e-05 ***
## Year2005 0.3060 0.0659 4.64 3.6e-06 ***
## Year2006 0.2763 0.0656 4.21 2.6e-05 ***
```

```

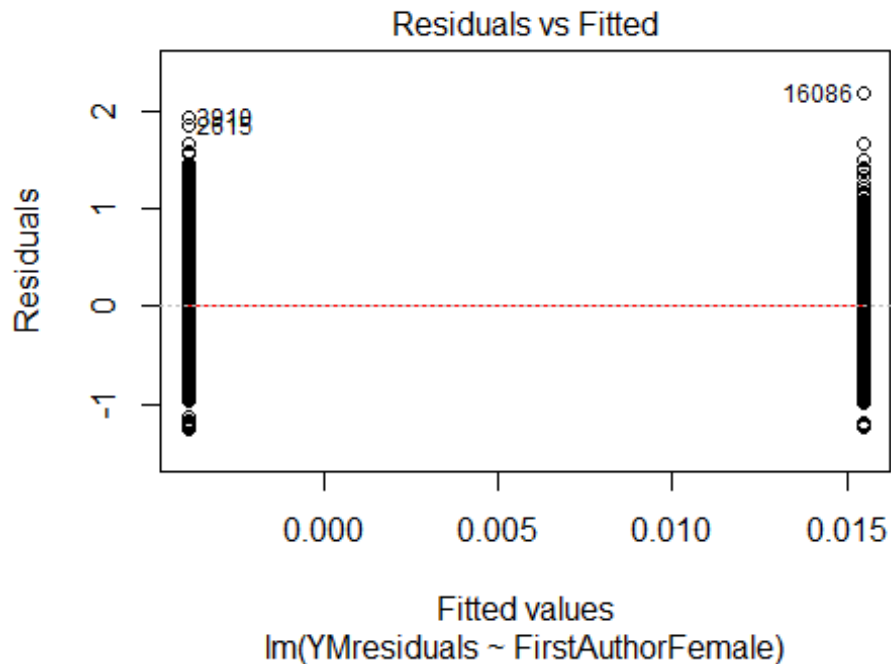
## Year2007          0.2823      0.0640      4.41  1.1e-05 ***
## Year2008          0.2756      0.0630      4.38  1.2e-05 ***
## Year2009          0.2506      0.0651      3.85  0.00012 ***
## Year2010          0.2444      0.0639      3.82  0.00013 ***
## Year2011          0.2899      0.0632      4.58  4.7e-06 ***
## Year2012          0.2256      0.0638      3.54  0.00041 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.0151
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 4023 is an outlier with |weight| = 0 ( < 2.2e-05);
## 380 weights are ~= 1. The remaining 4173 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0967 0.8590 0.9500 0.8920 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4554"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2505"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1380 1380 1350 1351 1274 1148 1037 927 905 848 928 811 776 822 846
## 2011 2012
## 895 749
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 597 601 661 568 442 365 545 499 501 477 473 415 432 408 436

```

```
## 2011 2012
## 499 393
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 528 515 581 509 387 322 484 445 438 423 426 369 385 365 395
## 2011 2012
## 444 344
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 37, df = 16, p-value = 0.002
```

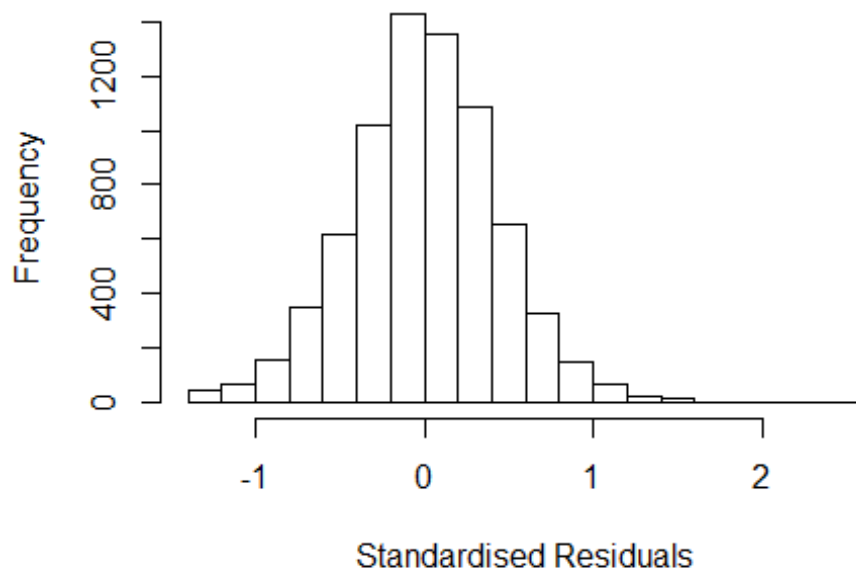


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 4.61250637852714"
## [1] "Male first author team size 2018 geometric mean: 3.67250378236879"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.54105080832256"
## [1] "Male last author team size 2018 geometric mean: 3.8218881180551"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.045 1 1.022
## LastAuthorFemale 1.016 1 1.008
## UniqueAuthors 1.063 4 1.008
## Year 1.116 16 1.003
```

## Residuals from first and last author and team size



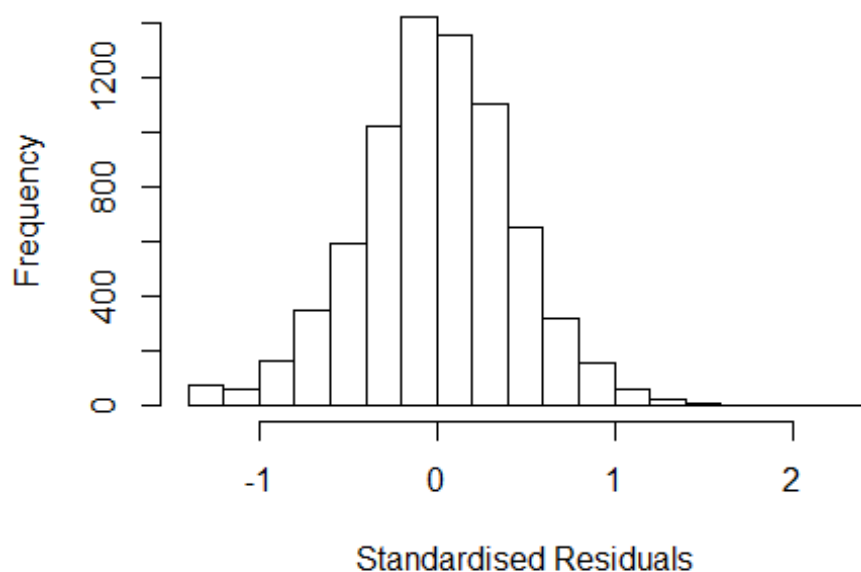
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27682 -0.27232 -0.00128 0.28112 2.40407
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.997848 0.036920 27.03 < 2e-16 ***
## FirstAuthorFemale1 0.013330 0.012237 1.09 0.2761
## LastAuthorFemale1 -0.049592 0.015867 -3.13 0.0018 **
## UniqueAuthors2 0.187499 0.034382 5.45 5.1e-08 ***
## UniqueAuthors3 0.177524 0.033724 5.26 1.4e-07 ***
## UniqueAuthors4 0.218976 0.033978 6.44 1.2e-10 ***
## UniqueAuthors5 0.219297 0.033503 6.55 6.3e-11 ***
## Year1997 0.040168 0.027769 1.45 0.1481
## Year1998 0.066716 0.027283 2.45 0.0145 *
## Year1999 -0.000953 0.027591 -0.03 0.9724
```

```

## Year2000      0.016413    0.029367    0.56    0.5762
## Year2001      0.048954    0.031193    1.57    0.1166
## Year2002      0.053529    0.027446    1.95    0.0512 .
## Year2003      0.005180    0.026680    0.19    0.8461
## Year2004      0.012887    0.027994    0.46    0.6453
## Year2005      0.036518    0.029263    1.25    0.2121
## Year2006      0.000855    0.027796    0.03    0.9755
## Year2007      0.046341    0.029744    1.56    0.1193
## Year2008      0.036254    0.029595    1.23    0.2206
## Year2009      0.036520    0.030641    1.19    0.2333
## Year2010     -0.018067    0.031799   -0.57    0.5699
## Year2011      0.032383    0.027940    1.16    0.2465
## Year2012     -0.036391    0.031060   -1.17    0.2414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0209, Adjusted R-squared:  0.0179
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 3 observations c(803,1266,5230) are outliers with |weight| = 0 ( < 1.4e-
05);
## 610 weights are ~= 1. The remaining 6747 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.066  0.866  0.951  0.898  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.36e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##           0          1000          0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1          1.021
## LastAuthorFemale  1.013 1          1.007
## Year              1.054 16          1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.255208 -0.273361 -0.000605 0.279396 2.205689
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18229 0.01948 60.70 <2e-16 ***
## FirstAuthorFemale1 0.01980 0.01225 1.62 0.1061
## LastAuthorFemale1 -0.04890 0.01587 -3.08 0.0021 **
## Year1997 0.03705 0.02786 1.33 0.1836
## Year1998 0.06935 0.02730 2.54 0.0111 *
## Year1999 -0.00015 0.02737 -0.01 0.9956
## Year2000 0.02271 0.02931 0.77 0.4385
## Year2001 0.04953 0.03105 1.60 0.1107
## Year2002 0.06119 0.02759 2.22 0.0266 *
## Year2003 0.01713 0.02660 0.64 0.5197
## Year2004 0.01704 0.02807 0.61 0.5438
## Year2005 0.04545 0.02931 1.55 0.1210
```

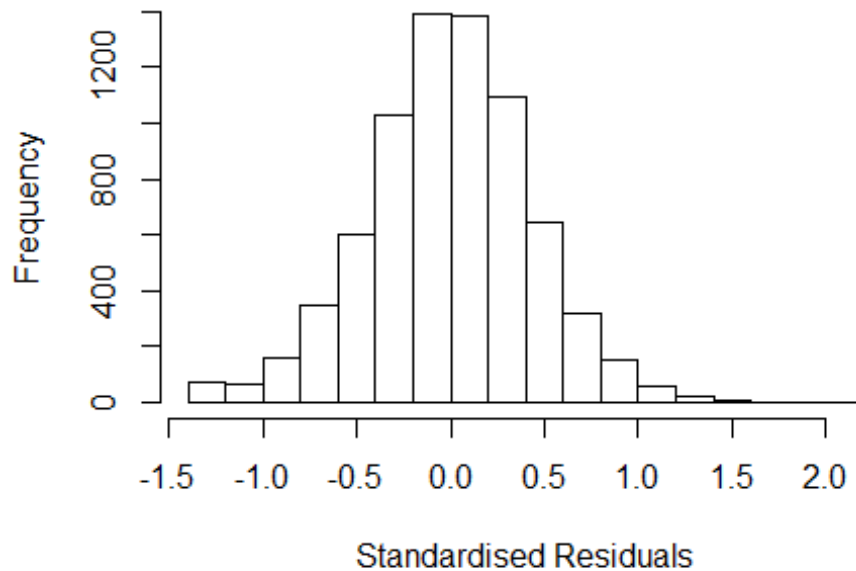
```

## Year2006          0.00661    0.02775    0.24    0.8118
## Year2007          0.05312    0.02991    1.78    0.0758 .
## Year2008          0.04589    0.02957    1.55    0.1207
## Year2009          0.04714    0.03081    1.53    0.1261
## Year2010         -0.00690    0.03150   -0.22    0.8266
## Year2011          0.04091    0.02783    1.47    0.1416
## Year2012         -0.02898    0.03117   -0.93    0.3524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.00555,    Adjusted R-squared:  0.00311
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 5230 is an outlier with |weight| = 0 ( < 1.4e-05);
## 626 weights are ~= 1. The remaining 6733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8650 0.9500 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1          1.021
## Year              1.042 16          1.001

```



## Residuals from first author



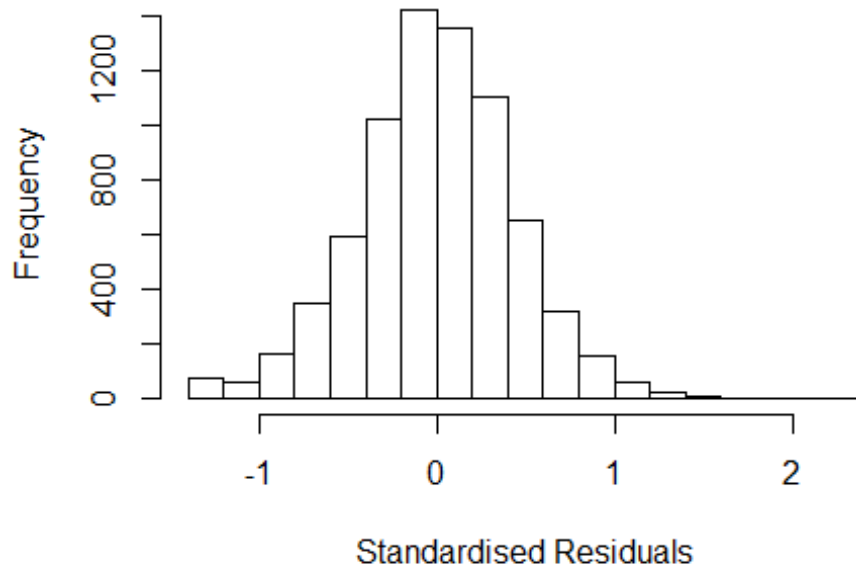
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.249703 -0.275357  0.000538  0.280106  2.162297
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.17790    0.01939   60.76  <2e-16 ***
## FirstAuthorFemale1  0.01766    0.01227    1.44   0.150
## Year1997          0.03676    0.02784    1.32   0.187
## Year1998          0.06926    0.02727    2.54   0.011 *
## Year1999         -0.00176    0.02737   -0.06   0.949
## Year2000          0.02056    0.02932    0.70   0.483
## Year2001          0.04985    0.03105    1.61   0.108
## Year2002          0.06000    0.02761    2.17   0.030 *
## Year2003          0.01625    0.02662    0.61   0.542
## Year2004          0.01662    0.02808    0.59   0.554
## Year2005          0.04526    0.02926    1.55   0.122
## Year2006          0.00597    0.02772    0.22   0.830
```

```

## Year2007          0.05415    0.02989    1.81    0.070 .
## Year2008          0.04592    0.02957    1.55    0.120
## Year2009          0.04656    0.03092    1.51    0.132
## Year2010         -0.00958    0.03139   -0.31    0.760
## Year2011          0.03865    0.02776    1.39    0.164
## Year2012         -0.03193    0.03112   -1.03    0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.00426,    Adjusted R-squared:  0.00196
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 5230 is an outlier with |weight| = 0 ( < 1.4e-05);
## 618 weights are ~= 1. The remaining 6741 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  0.865   0.950   0.896   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year            1.012 16          1.000

```

## Residuals from last author



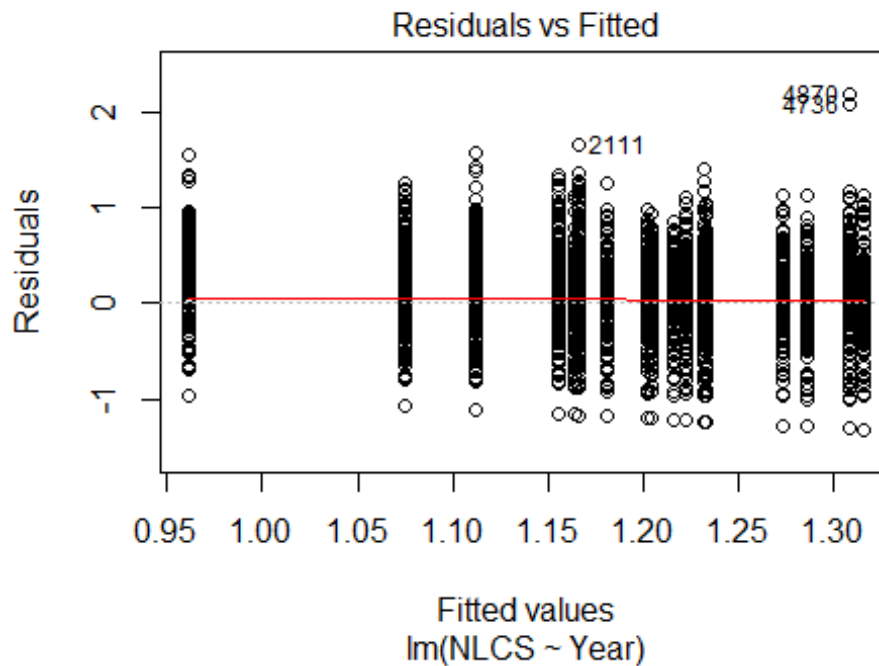
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.255067 -0.272267 -0.000756 0.278917 2.219594
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.184474 0.019419 61.00 <2e-16 ***
## LastAuthorFemale1 -0.047490 0.015853 -3.00 0.0027 **
## Year1997 0.037887 0.027833 1.36 0.1735
## Year1998 0.070593 0.027256 2.59 0.0096 **
## Year1999 0.000353 0.027350 0.01 0.9897
## Year2000 0.023383 0.029316 0.80 0.4251
## Year2001 0.050503 0.031054 1.63 0.1039
## Year2002 0.062615 0.027525 2.27 0.0229 *
## Year2003 0.019170 0.026528 0.72 0.4699
## Year2004 0.018491 0.028045 0.66 0.5097
## Year2005 0.047771 0.029204 1.64 0.1019
## Year2006 0.009472 0.027646 0.34 0.7319
```

```

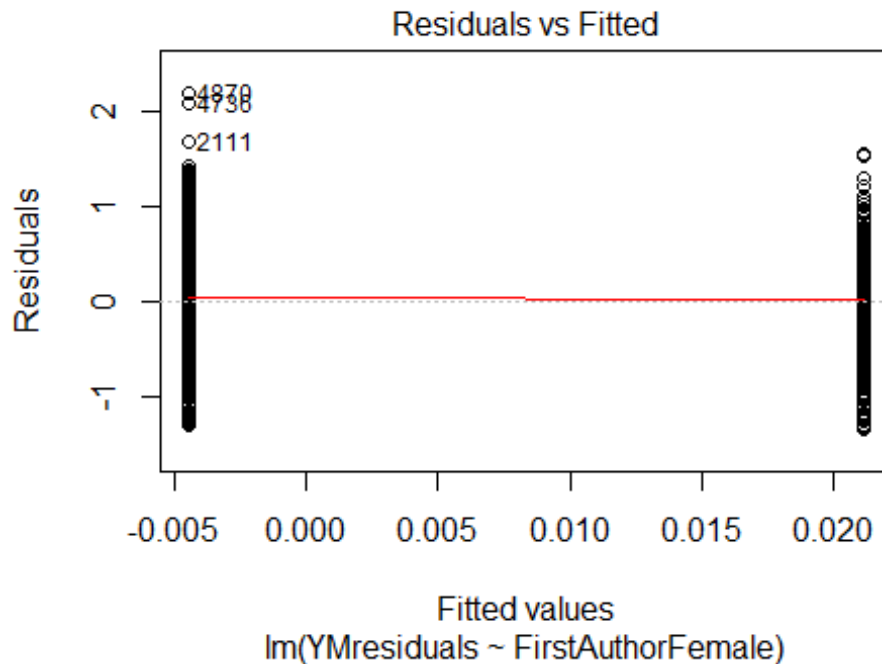
## Year2007      0.055423    0.029896    1.85    0.0638 .
## Year2008      0.048384    0.029465    1.64    0.1006
## Year2009      0.049830    0.030730    1.62    0.1049
## Year2010     -0.003637    0.031270   -0.12    0.9074
## Year2011      0.044886    0.027578    1.63    0.1036
## Year2012     -0.025641    0.031018   -0.83    0.4085
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.00521,    Adjusted R-squared:  0.00291
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 5230 is an outlier with |weight| = 0 ( < 1.4e-05);
## 615 weights are ~= 1. The remaining 6744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8660 0.9500 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7360"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2506"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 683 727 739 663 663 527 482 428 425 360 408 333 345 368 343
## 2011 2012
## 397 385
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 325 298 341 253 202 124 242 209 194 172 176 161 179 177 172

```

```
## 2011 2012
## 216 202
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 284 262 312 230 183 103 216 185 177 151 159 138 162 160 159
## 2011 2012
## 191 185
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```

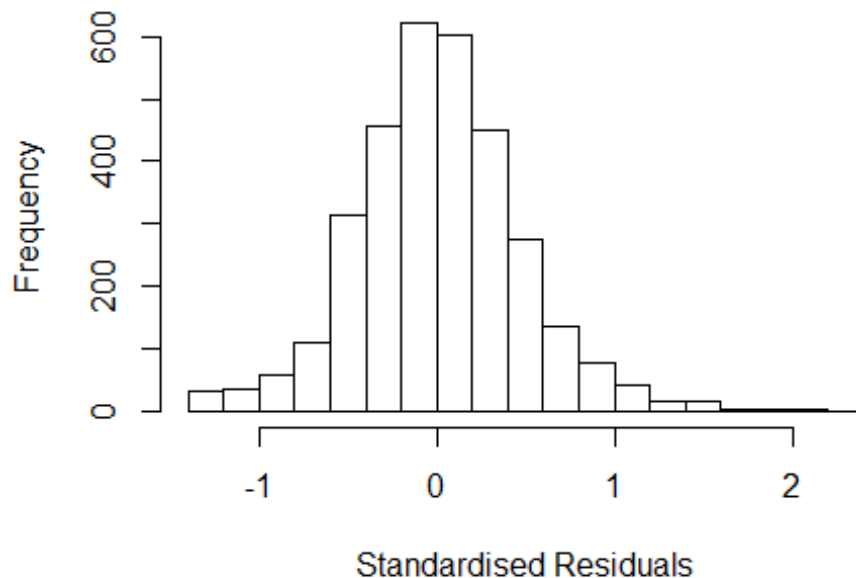


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.57113531305214"
## [1] "Male first author team size 2018 geometric mean: 3.70337849732955"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.50802441269159"
## [1] "Male last author team size 2018 geometric mean: 3.83145999072598"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1          1.029
## LastAuthorFemale  1.037 1          1.018
## UniqueAuthors    1.253 4          1.029
## Year             1.323 16          1.009
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.365587 -0.267941 -0.000556 0.282242 2.212880
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.493547 0.045767 10.78 < 2e-16 ***
## FirstAuthorFemale1 -0.029008 0.019165 -1.51 0.13023
## LastAuthorFemale1 -0.033649 0.024714 -1.36 0.17344
## UniqueAuthors2 0.672868 0.045566 14.77 < 2e-16 ***
## UniqueAuthors3 0.724395 0.041250 17.56 < 2e-16 ***
## UniqueAuthors4 0.798770 0.040940 19.51 < 2e-16 ***
## UniqueAuthors5 0.778145 0.040768 19.09 < 2e-16 ***
## Year1997 0.009523 0.042865 0.22 0.82420
## Year1998 0.053347 0.041791 1.28 0.20186
## Year1999 0.000258 0.043193 0.01 0.99524
```

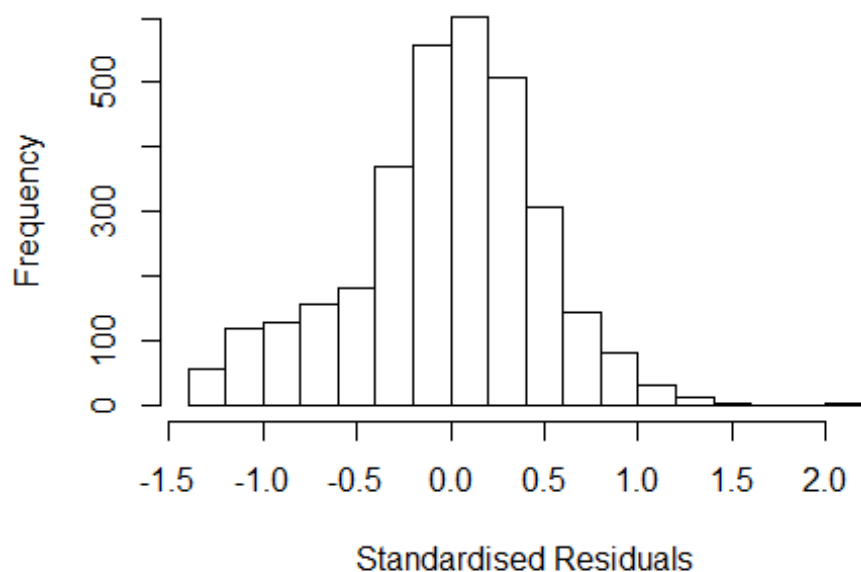
```

## Year2000      -0.125770    0.046641    -2.70    0.00704 **
## Year2001      0.181359    0.058058     3.12    0.00180 **
## Year2002      0.144424    0.041162     3.51    0.00046 ***
## Year2003      0.048449    0.040824     1.19    0.23540
## Year2004      0.073270    0.044059     1.66    0.09641 .
## Year2005      0.035622    0.045113     0.79    0.42980
## Year2006      0.047923    0.045237     1.06    0.28950
## Year2007      0.040290    0.045304     0.89    0.37390
## Year2008      0.065386    0.043100     1.52    0.12935
## Year2009      0.083886    0.042420     1.98    0.04807 *
## Year2010      0.034637    0.045513     0.76    0.44669
## Year2011      0.098929    0.041883     2.36    0.01823 *
## Year2012     -0.032284    0.040529    -0.80    0.42576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.275, Adjusted R-squared:  0.27
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 7 observations c(262,266,340,595,1153,1407,1434)
## are outliers with |weight| = 0 ( < 3.1e-05);
## 272 weights are ~ = 1. The remaining 2978 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0018 0.8590 0.9460 0.8820 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.07e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.076 1 1.037
## LastAuthorFemale 1.051 1 1.025
## Year 1.081 16 1.002

```



## Residuals from first and last author



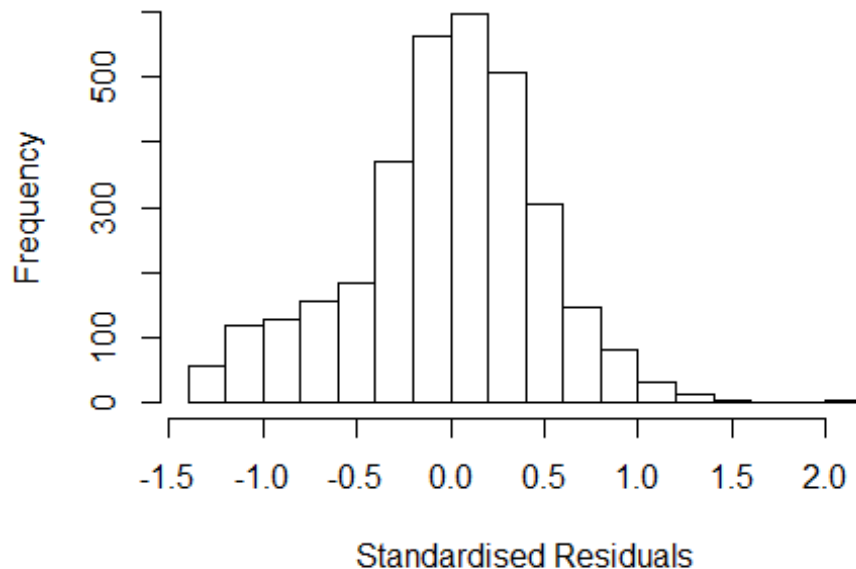
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3676 -0.2888 0.0194 0.2984 2.1753
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.134813 0.035865 31.64 < 2e-16 ***
## FirstAuthorFemale1 0.022818 0.020815 1.10 0.27308
## LastAuthorFemale1 0.000736 0.027382 0.03 0.97856
## Year1997 0.039082 0.051535 0.76 0.44830
## Year1998 0.053742 0.049965 1.08 0.28219
## Year1999 -0.027245 0.054255 -0.50 0.61558
## Year2000 -0.178141 0.066668 -2.67 0.00758 **
## Year2001 0.209237 0.057849 3.62 0.00030 ***
## Year2002 0.176893 0.047726 3.71 0.00021 ***
## Year2003 0.106618 0.044479 2.40 0.01658 *
## Year2004 0.094645 0.049628 1.91 0.05660 .
## Year2005 0.073333 0.053854 1.36 0.17339
```

```

## Year2006          0.071387    0.051199    1.39  0.16333
## Year2007          0.096054    0.051525    1.86  0.06238 .
## Year2008          0.132179    0.047974    2.76  0.00590 **
## Year2009          0.157558    0.048272    3.26  0.00111 **
## Year2010          0.095130    0.049930    1.91  0.05683 .
## Year2011          0.152159    0.046517    3.27  0.00108 **
## Year2012          0.039879    0.046804    0.85  0.39425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.025
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1407,1434) are outliers with |weight| = 0 ( < 3.1e-05);
## 283 weights are ~ = 1. The remaining 2972 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.135  0.840  0.950   0.881  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1          1.026
## Year              1.053 16          1.002

```

## Residuals from first author



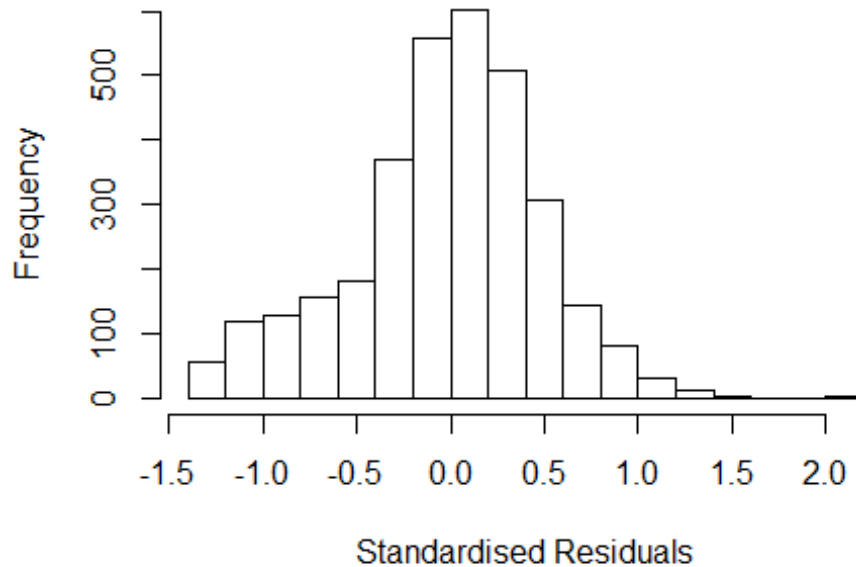
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3670 -0.2889 0.0193 0.2983 2.1752
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1349 0.0357 31.77 < 2e-16 ***
## FirstAuthorFemale1 0.0228 0.0209 1.09 0.27546
## Year1997 0.0391 0.0515 0.76 0.44838
## Year1998 0.0537 0.0500 1.08 0.28218
## Year1999 -0.0272 0.0542 -0.50 0.61572
## Year2000 -0.1782 0.0666 -2.67 0.00754 **
## Year2001 0.2093 0.0578 3.62 0.00030 ***
## Year2002 0.1769 0.0477 3.71 0.00021 ***
## Year2003 0.1066 0.0445 2.40 0.01660 *
## Year2004 0.0946 0.0496 1.91 0.05664 .
## Year2005 0.0733 0.0539 1.36 0.17343
## Year2006 0.0714 0.0512 1.39 0.16327
```

```

## Year2007          0.0960      0.0515      1.86  0.06245 .
## Year2008          0.1322      0.0480      2.76  0.00590 **
## Year2009          0.1575      0.0483      3.26  0.00111 **
## Year2010          0.0951      0.0499      1.91  0.05665 .
## Year2011          0.1522      0.0464      3.28  0.00105 **
## Year2012          0.0399      0.0468      0.85  0.39399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0304, Adjusted R-squared:  0.0253
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1407,1434) are outliers with |weight| = 0 ( < 3.1e-05);
## 282 weights are ~= 1. The remaining 2973 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.135  0.840  0.950  0.881  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.025 1          1.013
## Year          1.025 16          1.001

```

## Residuals from last author



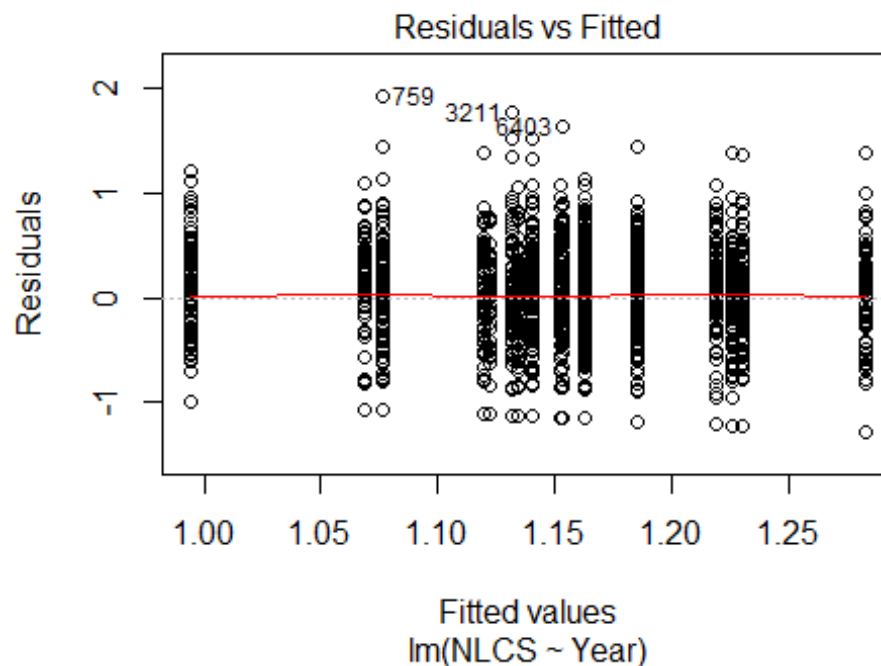
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3479 -0.2906 0.0219 0.2971 2.1718
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13738 0.03567 31.88 < 2e-16 ***
## LastAuthorFemale1 0.00166 0.02745 0.06 0.95173
## Year1997 0.03969 0.05152 0.77 0.44112
## Year1998 0.05524 0.04991 1.11 0.26849
## Year1999 -0.02761 0.05428 -0.51 0.61097
## Year2000 -0.17878 0.06676 -2.68 0.00744 **
## Year2001 0.20885 0.05782 3.61 0.00031 ***
## Year2002 0.17781 0.04767 3.73 0.00019 ***
## Year2003 0.10788 0.04447 2.43 0.01532 *
## Year2004 0.09600 0.04962 1.93 0.05311 .
## Year2005 0.07546 0.05382 1.40 0.16100
## Year2006 0.07431 0.05121 1.45 0.14687
```

```

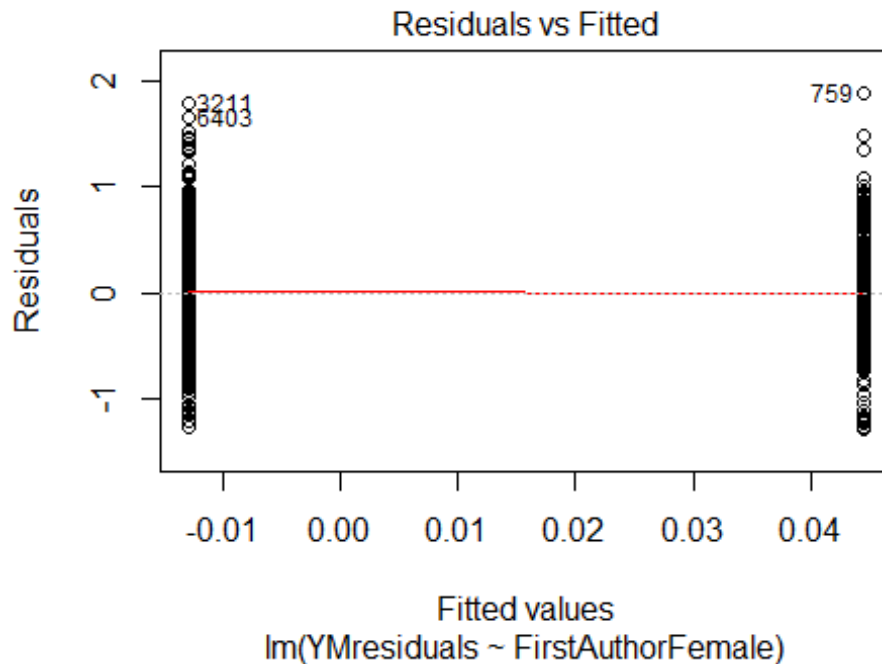
## Year2007      0.09789      0.05151      1.90  0.05746 .
## Year2008      0.13556      0.04772      2.84  0.00453 **
## Year2009      0.16014      0.04825      3.32  0.00091 ***
## Year2010      0.09821      0.04975      1.97  0.04846 *
## Year2011      0.15799      0.04603      3.43  0.00061 ***
## Year2012      0.04415      0.04659      0.95  0.34337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.44
## Multiple R-squared:  0.0302, Adjusted R-squared:  0.0251
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1407,1434) are outliers with |weight| = 0 ( < 3.1e-05);
## 287 weights are ~= 1. The remaining 2968 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.137  0.840   0.949   0.880   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3257"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2507"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   504  473  442  424  423  328  287  260  241  226  267  227  213  191  212
## 2011 2012
##   254  239
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   131  119  134   96   88   53   87   75   82   75  101   91  108   75   98

```

```
## 2011 2012
## 131 104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 112 104 120 90 78 48 76 68 75 61 85 71 97 66 85
## 2011 2012
## 110 82
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```



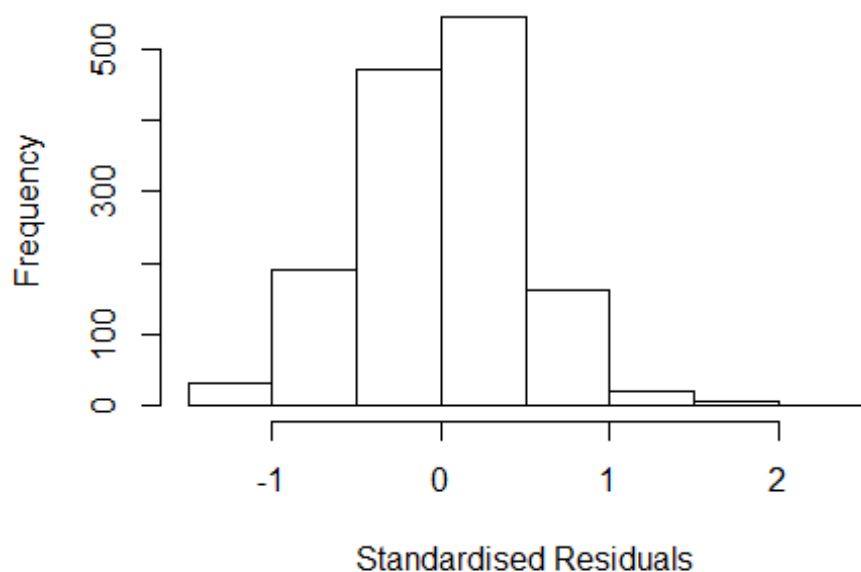
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.5, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 3.81045160569668"
## [1] "Male first author team size 2018 geometric mean: 3.68571310308385"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 900, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.11284261365502"
## [1] "Male last author team size 2018 geometric mean: 3.86566561549305"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.125 1      1.061
## LastAuthorFemale  1.099 1      1.048
## UniqueAuthors    1.296 4      1.033
## Year             1.460 16      1.012
```



## Residuals from first and last author and team size



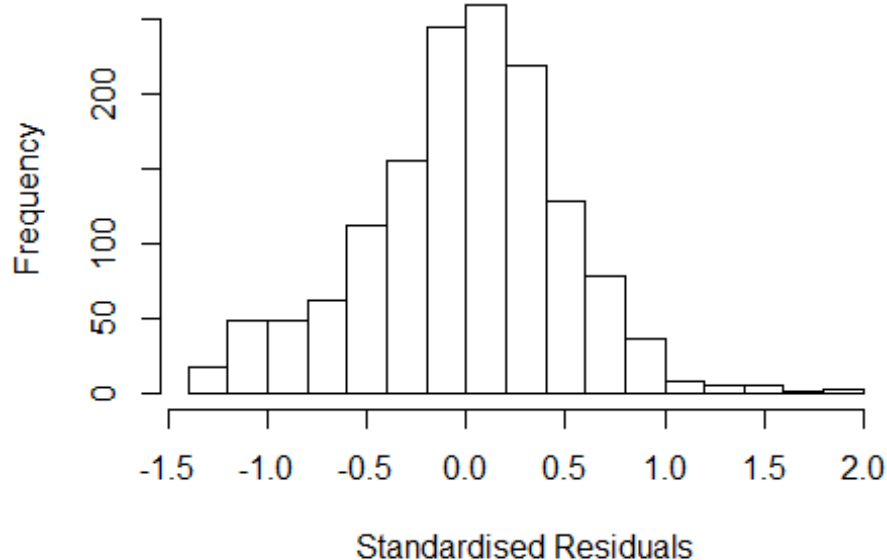
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4217 -0.2845 0.0162 0.3067 2.2831
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.85542 0.06308 13.56 < 2e-16 ***
## FirstAuthorFemale1 0.01561 0.02917 0.54 0.5925
## LastAuthorFemale1 -0.07632 0.04171 -1.83 0.0675 .
## UniqueAuthors2 0.37131 0.05494 6.76 2.0e-11 ***
## UniqueAuthors3 0.41867 0.05397 7.76 1.7e-14 ***
## UniqueAuthors4 0.42787 0.05674 7.54 8.4e-14 ***
## UniqueAuthors5 0.49303 0.05605 8.80 < 2e-16 ***
## Year1997 -0.07284 0.07308 -1.00 0.3191
## Year1998 -0.05061 0.06646 -0.76 0.4465
## Year1999 -0.22331 0.07122 -3.14 0.0018 **
```

```

## Year2000      0.05692      0.07905      0.72      0.4716
## Year2001     -0.03129      0.08790     -0.36      0.7219
## Year2002     -0.11721      0.08348     -1.40      0.1605
## Year2003     -0.14388      0.07215     -1.99      0.0463 *
## Year2004     -0.05331      0.07376     -0.72      0.4700
## Year2005      0.13841      0.07393      1.87      0.0614 .
## Year2006     -0.17650      0.06286     -2.81      0.0051 **
## Year2007     -0.08649      0.07187     -1.20      0.2290
## Year2008      0.00126      0.06463      0.02      0.9844
## Year2009     -0.05175      0.07533     -0.69      0.4922
## Year2010     -0.01950      0.07065     -0.28      0.7826
## Year2011     -0.08281      0.06492     -1.28      0.2023
## Year2012     -0.05983      0.07330     -0.82      0.4145
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.451
## Multiple R-squared:  0.122, Adjusted R-squared:  0.108
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 141 is an outlier with |weight| = 0 ( < 7e-05);
## 120 weights are ~= 1. The remaining 1307 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0335 0.8590 0.9510 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.00e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max      maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.089 1      1.044
## LastAuthorFemale 1.089 1      1.043
## Year      1.169 16      1.005

```

## Residuals from first and last author



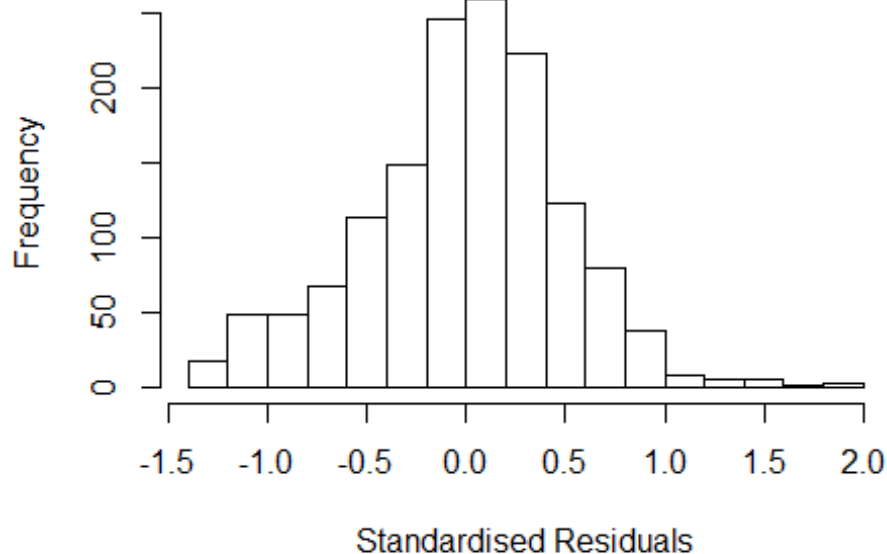
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3422 -0.3003 0.0238 0.3163 1.9274
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18390 0.04891 24.21 <2e-16 ***
## FirstAuthorFemale1 0.04780 0.02997 1.59 0.1110
## LastAuthorFemale1 -0.06341 0.04288 -1.48 0.1394
## Year1997 -0.09071 0.07671 -1.18 0.2372
## Year1998 -0.04769 0.07084 -0.67 0.5009
## Year1999 -0.20747 0.07258 -2.86 0.0043 **
## Year2000 0.07994 0.08144 0.98 0.3265
## Year2001 -0.01331 0.08770 -0.15 0.8794
## Year2002 -0.10977 0.08568 -1.28 0.2003
## Year2003 -0.09891 0.07936 -1.25 0.2128
## Year2004 -0.02781 0.07751 -0.36 0.7198
## Year2005 0.15826 0.07460 2.12 0.0341 *
```

```

## Year2006      -0.09781    0.06400   -1.53    0.1267
## Year2007      -0.01480    0.07448   -0.20    0.8425
## Year2008       0.07368    0.06711    1.10    0.2724
## Year2009       0.03031    0.07630    0.40    0.6913
## Year2010       0.06616    0.06896    0.96    0.3375
## Year2011      -0.00888    0.06605   -0.13    0.8931
## Year2012      -0.01763    0.07264   -0.24    0.8083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.0325, Adjusted R-squared:  0.0202
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 118 weights are ~= 1. The remaining 1310 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.035  0.856  0.949  0.891  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.00e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.088 1      1.043
## Year              1.088 16      1.003

```

## Residuals from first author



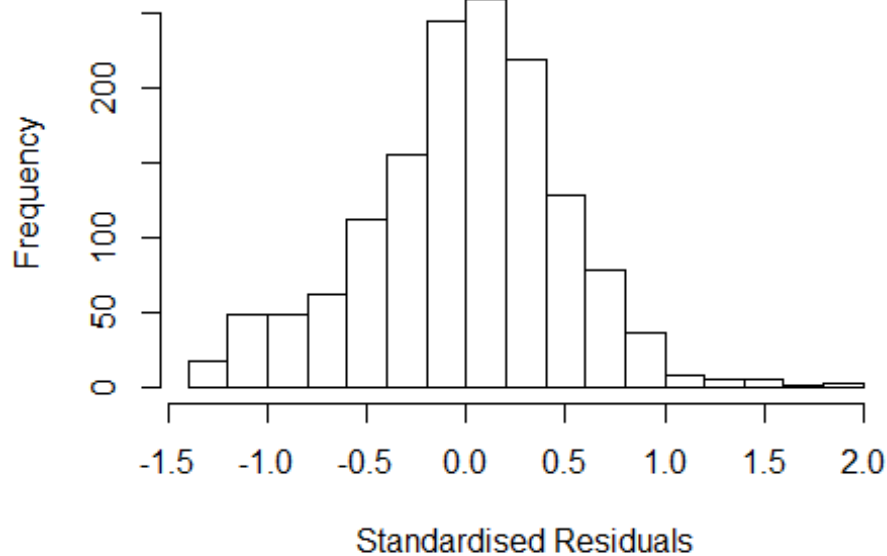
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3375 -0.3068  0.0247  0.3154  1.8752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1780     0.0486   24.21  <2e-16 ***
## FirstAuthorFemale1  0.0421     0.0301    1.40   0.1628
## Year1997         -0.0903     0.0764   -1.18   0.2375
## Year1998         -0.0500     0.0709   -0.70   0.4810
## Year1999         -0.2086     0.0727   -2.87   0.0042 **
## Year2000          0.0818     0.0812    1.01   0.3139
## Year2001         -0.0129     0.0883   -0.15   0.8839
## Year2002         -0.1107     0.0857   -1.29   0.1964
## Year2003         -0.1029     0.0796   -1.29   0.1960
## Year2004         -0.0265     0.0780   -0.34   0.7337
## Year2005          0.1595     0.0743    2.15   0.0320 *
## Year2006         -0.0962     0.0638   -1.51   0.1321
```

```

## Year2007          -0.0105      0.0745   -0.14   0.8880
## Year2008           0.0754      0.0669    1.13   0.2595
## Year2009           0.0248      0.0764    0.33   0.7452
## Year2010           0.0651      0.0682    0.96   0.3396
## Year2011          -0.0105      0.0656   -0.16   0.8733
## Year2012          -0.0216      0.0714   -0.30   0.7627
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.031, Adjusted R-squared:  0.0193
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1304 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0534 0.8560 0.9490 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.00e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.079 1          1.039
## Year            1.079 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3527 -0.3044  0.0289  0.3143  1.9620
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.189721    0.048577   24.49  <2e-16 ***
## LastAuthorFemale1 -0.055928    0.042388   -1.32   0.1872
## Year1997        -0.090783    0.076633   -1.18   0.2364
## Year1998        -0.044180    0.070898   -0.62   0.5333
## Year1999        -0.208627    0.072618   -2.87   0.0041 **
## Year2000         0.083752    0.080749    1.04   0.2998
## Year2001        -0.012756    0.087807   -0.15   0.8845
## Year2002        -0.112979    0.085456   -1.32   0.1864
## Year2003        -0.089808    0.078679   -1.14   0.2539
## Year2004        -0.022738    0.078282   -0.29   0.7715
## Year2005         0.162942    0.074351    2.19   0.0286 *
## Year2006        -0.087836    0.063591   -1.38   0.1674
```

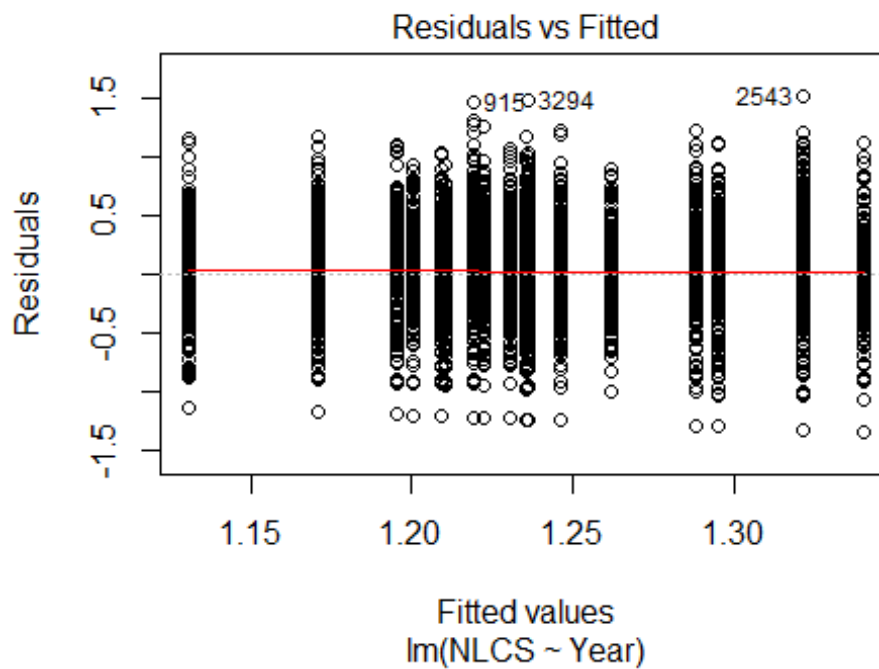
```

## Year2007          -0.006778    0.074500   -0.09    0.9275
## Year2008           0.078739    0.066849    1.18    0.2391
## Year2009           0.040329    0.075965    0.53    0.5956
## Year2010           0.075650    0.068016    1.11    0.2662
## Year2011          -0.000218    0.065467    0.00    0.9973
## Year2012          -0.007867    0.072253   -0.11    0.9133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.0311, Adjusted R-squared:  0.0195
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1316 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0245 0.8560 0.9490 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.00e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1428"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2508"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 807 852 825 791 753 707 601 556 524 490 619 591 588 656 590
## 2011 2012
## 609 532
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 309 327 351 272 208 178 293 258 246 228 244 265 277 285 267
## 2011 2012

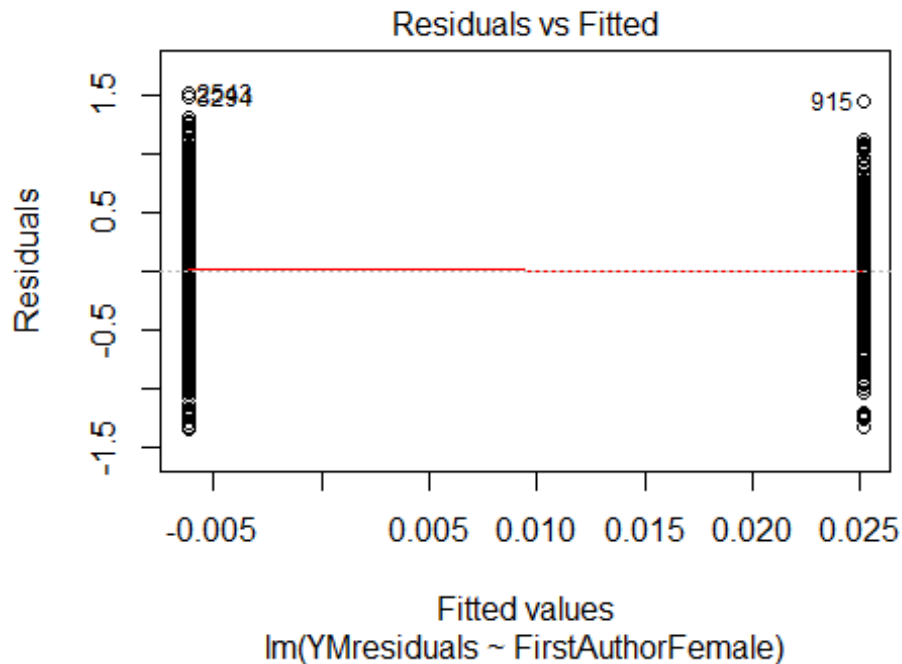
```



```
## 313 293
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 265 277 314 245 186 156 262 232 217 204 216 229 248 256 248
## 2011 2012
## 274 256
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 61, df = 16, p-value = 4e-07
```

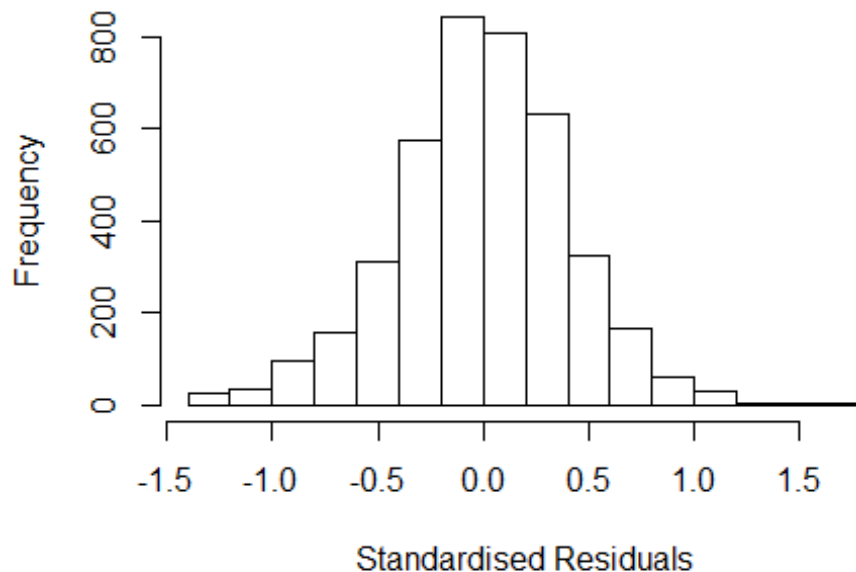


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 22, df = 1, p-value = 3e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.65889133919115"
## [1] "Male first author team size 2018 geometric mean: 3.82292599843565"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5200, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.49383979318327"
## [1] "Male last author team size 2018 geometric mean: 3.91270566743171"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3300, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.040 1      1.020
## LastAuthorFemale  1.032 1      1.016
## UniqueAuthors    1.093 4      1.011
## Year             1.136 16      1.004
```

## Residuals from first and last author and team size



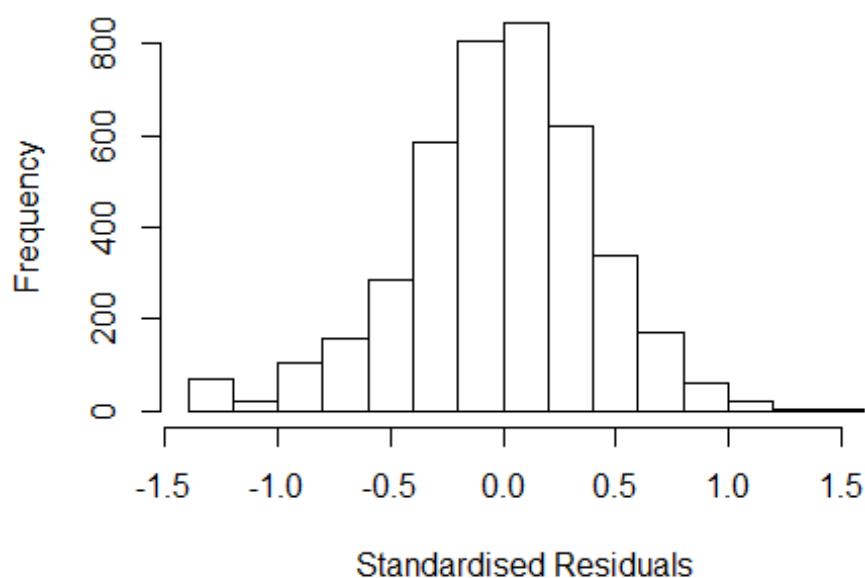
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.337246 -0.255204 -0.000868 0.254858 1.770907
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.21e-01 5.21e-02 17.67 < 2e-16 ***
## FirstAuthorFemale1 -1.27e-03 1.51e-02 -0.08 0.9330
## LastAuthorFemale1 -9.36e-05 2.00e-02 0.00 0.9963
## UniqueAuthors2 3.06e-01 4.76e-02 6.44 1.4e-10 ***
## UniqueAuthors3 3.12e-01 4.68e-02 6.67 2.9e-11 ***
## UniqueAuthors4 3.70e-01 4.71e-02 7.86 4.9e-15 ***
## UniqueAuthors5 3.68e-01 4.64e-02 7.94 2.7e-15 ***
## Year1997 4.40e-02 3.78e-02 1.16 0.2451
## Year1998 9.81e-02 3.81e-02 2.57 0.0102 *
## Year1999 1.50e-02 3.90e-02 0.39 0.7001
```

```

## Year2000      -1.99e-02  4.09e-02  -0.49  0.6269
## Year2001      1.10e-01  4.10e-02   2.68  0.0073 **
## Year2002      8.19e-02  3.75e-02   2.18  0.0290 *
## Year2003     -2.74e-02  3.65e-02  -0.75  0.4524
## Year2004      1.58e-02  3.73e-02   0.42  0.6717
## Year2005      1.61e-03  3.71e-02   0.04  0.9654
## Year2006     -1.47e-02  3.78e-02  -0.39  0.6970
## Year2007     -2.50e-03  3.73e-02  -0.07  0.9466
## Year2008     -1.74e-02  3.63e-02  -0.48  0.6326
## Year2009     -3.25e-04  3.65e-02  -0.01  0.9929
## Year2010     -6.17e-02  3.90e-02  -1.58  0.1133
## Year2011      1.62e-02  3.58e-02   0.45  0.6509
## Year2012     -9.05e-02  3.83e-02  -2.37  0.0180 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.379
## Multiple R-squared:  0.0627, Adjusted R-squared:  0.0576
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 350 weights are ~= 1. The remaining 3735 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000  0.868  0.950  0.892  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.45e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1      1.015
## LastAuthorFemale  1.020 1      1.010
## Year              1.051 16      1.002

```

## Residuals from first and last author



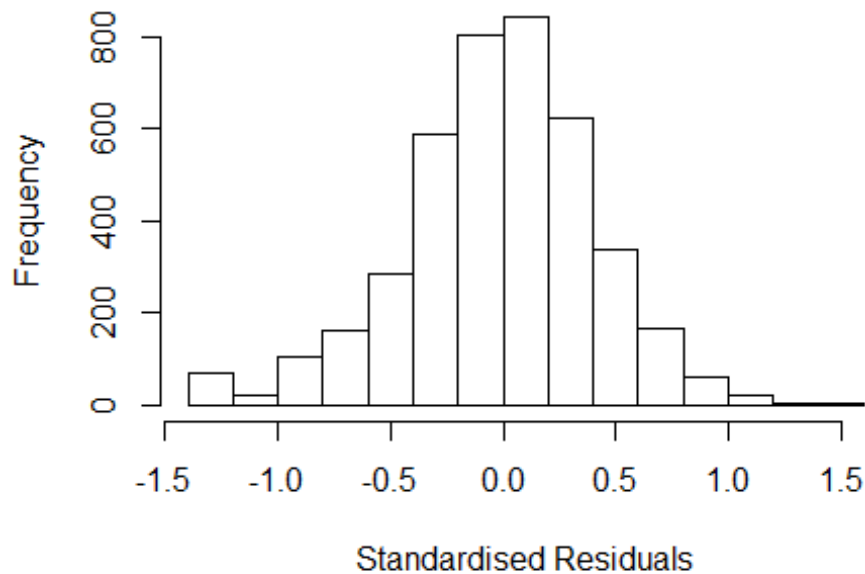
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35143 -0.25882  0.00441  0.25558  1.49186
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.23783    0.02728   45.37  <2e-16 ***
## FirstAuthorFemale1  0.01383    0.01508    0.92   0.3591
## LastAuthorFemale1  0.00703    0.01998    0.35   0.7248
## Year1997         0.03923    0.03813    1.03   0.3036
## Year1998         0.09931    0.03791    2.62   0.0088 **
## Year1999         0.00254    0.03949    0.06   0.9486
## Year2000        -0.01791    0.04139   -0.43   0.6652
## Year2001         0.10656    0.04082    2.61   0.0091 **
## Year2002         0.07980    0.03741    2.13   0.0330 *
## Year2003        -0.01424    0.03616   -0.39   0.6938
## Year2004         0.01584    0.03737    0.42   0.6717
## Year2005         0.01236    0.03732    0.33   0.7406
```

```

## Year2006      -0.01035    0.03803   -0.27    0.7855
## Year2007      0.01099    0.03741    0.29    0.7688
## Year2008     -0.00520    0.03639   -0.14    0.8863
## Year2009      0.01243    0.03703    0.34    0.7372
## Year2010     -0.05451    0.03899   -1.40    0.1621
## Year2011      0.01857    0.03571    0.52    0.6031
## Year2012     -0.08324    0.03817   -2.18    0.0293 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.014, Adjusted R-squared:  0.00966
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 355 weights are ~= 1. The remaining 3730 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0923 0.8670 0.9500 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.45e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.03 1      1.015
## Year      1.03 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34528 -0.25684 0.00363 0.25519 1.49119
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23847 0.02719 45.54 <2e-16 ***
## FirstAuthorFemale1 0.01405 0.01509 0.93 0.3519
## Year1997 0.03925 0.03813 1.03 0.3034
## Year1998 0.09934 0.03791 2.62 0.0088 **
## Year1999 0.00272 0.03945 0.07 0.9449
## Year2000 -0.01783 0.04140 -0.43 0.6667
## Year2001 0.10681 0.04080 2.62 0.0089 **
## Year2002 0.07986 0.03740 2.14 0.0328 *
## Year2003 -0.01408 0.03615 -0.39 0.6969
## Year2004 0.01569 0.03736 0.42 0.6745
## Year2005 0.01215 0.03731 0.33 0.7448
## Year2006 -0.01022 0.03802 -0.27 0.7882
```

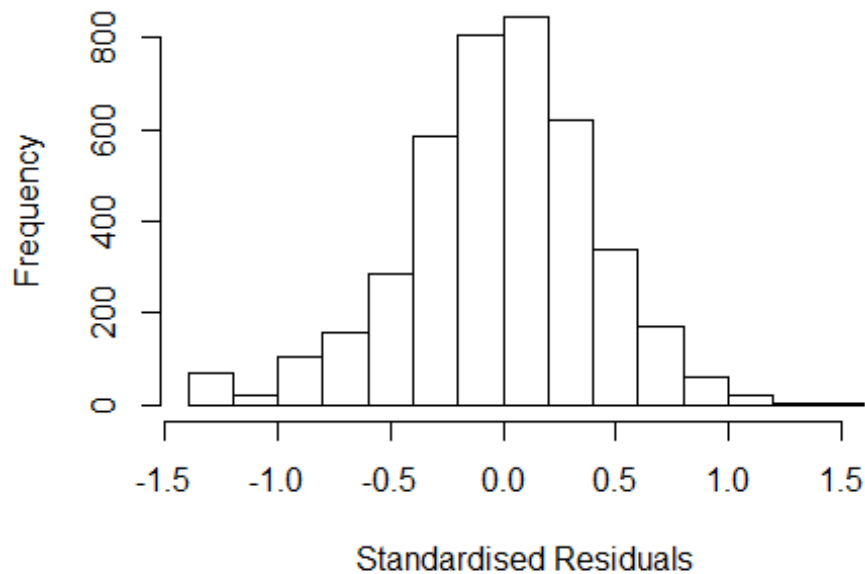
```

## Year2007          0.01097    0.03741    0.29    0.7694
## Year2008         -0.00531    0.03637   -0.15    0.8840
## Year2009          0.01232    0.03703    0.33    0.7394
## Year2010         -0.05433    0.03899   -1.39    0.1635
## Year2011          0.01885    0.03570    0.53    0.5975
## Year2012         -0.08293    0.03813   -2.18    0.0297 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.014, Adjusted R-squared:  0.00987
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 354 weights are ~= 1. The remaining 3731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0924 0.8670 0.9500 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.45e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1          1.010
## Year              1.021 16          1.001

```



## Residuals from last author



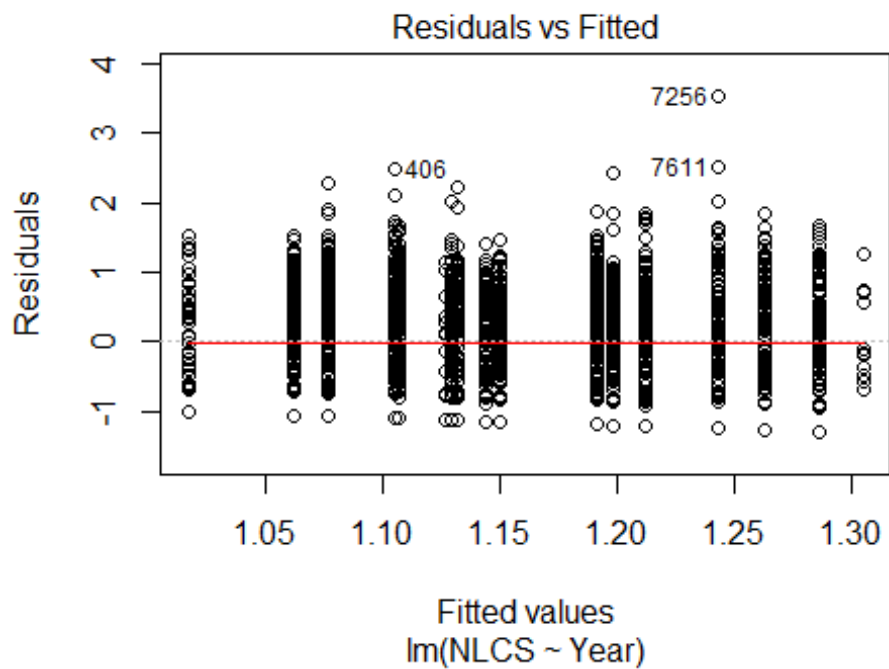
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35526 -0.25804 0.00513 0.25502 1.48918
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23945 0.02717 45.61 <2e-16 ***
## LastAuthorFemale1 0.00783 0.01997 0.39 0.695
## Year1997 0.04021 0.03805 1.06 0.291
## Year1998 0.10037 0.03785 2.65 0.008 **
## Year1999 0.00242 0.03948 0.06 0.951
## Year2000 -0.01740 0.04142 -0.42 0.674
## Year2001 0.10798 0.04072 2.65 0.008 **
## Year2002 0.08032 0.03737 2.15 0.032 *
## Year2003 -0.01312 0.03610 -0.36 0.716
## Year2004 0.01641 0.03736 0.44 0.661
## Year2005 0.01345 0.03729 0.36 0.718
## Year2006 -0.00848 0.03791 -0.22 0.823
```

```

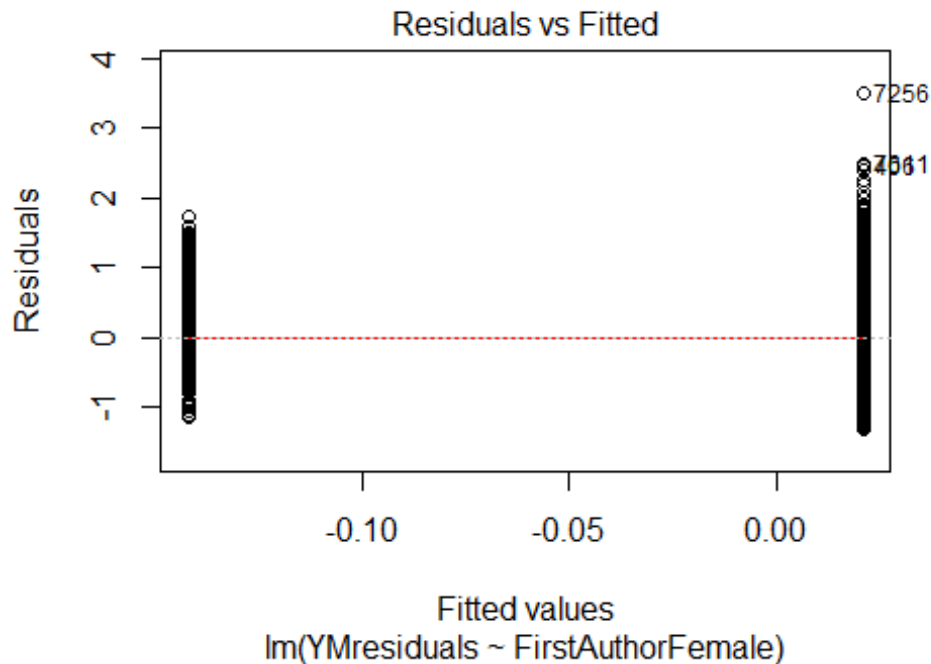
## Year2007      0.01250      0.03737      0.33      0.738
## Year2008     -0.00341      0.03625     -0.09      0.925
## Year2009      0.01427      0.03695      0.39      0.699
## Year2010     -0.05252      0.03888     -1.35      0.177
## Year2011      0.02140      0.03551      0.60      0.547
## Year2012     -0.08113      0.03801     -2.13      0.033 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.0138, Adjusted R-squared:  0.00973
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 355 weights are ~= 1. The remaining 3730 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0936 0.8670 0.9500 0.8890 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.45e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4085"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2600"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 458 173 42 472 22 315 261 363 300 289 388 492 465 491 406
## 2011 2012
## 428 419
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 195 56 16 203 14 130 129 197 157 160 238 312 273 308 260
## 2011 2012

```

```
## 272 268
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 185 52 16 191 13 126 120 186 152 155 219 287 254 288 235
## 2011 2012
## 251 255
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 6e-04
```

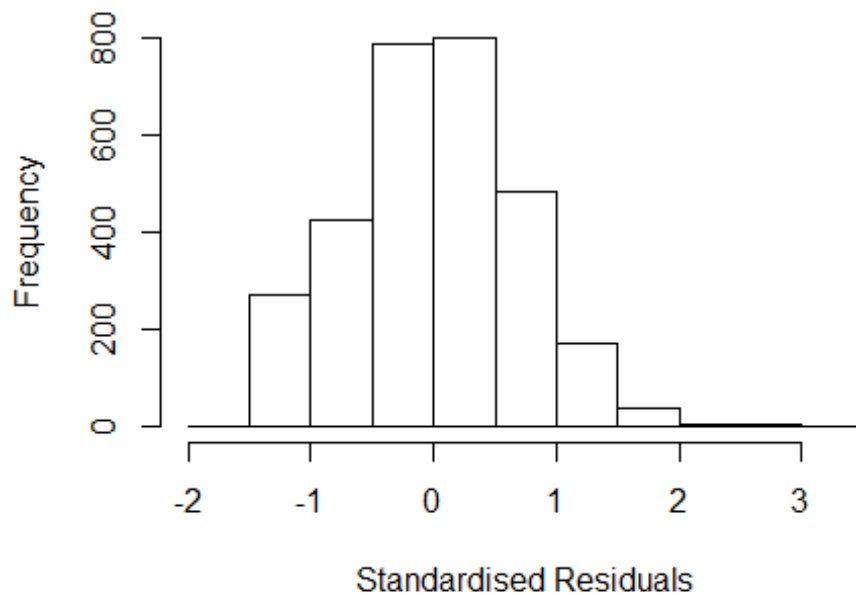


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 1.72044155330614"
## [1] "Male first author team size 2018 geometric mean: 1.47346899123064"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.68185214173174"
## [1] "Male last author team size 2018 geometric mean: 1.47927489449513"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.773 1 1.331
## LastAuthorFemale 1.763 1 1.328
## UniqueAuthors 1.127 4 1.015
## Year 1.163 16 1.005
```

## Residuals from first and last author and team size



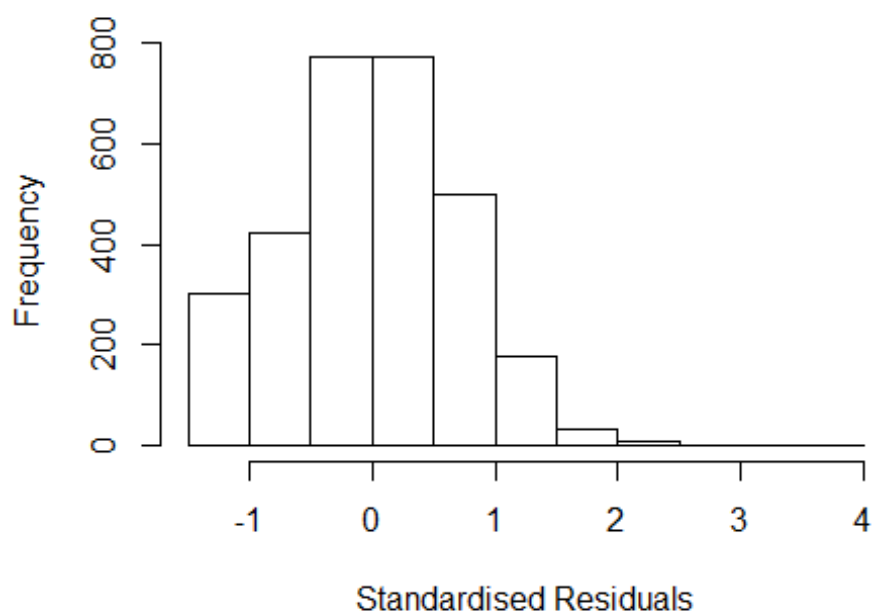
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 406   0007244732 3.574 1996    2200     3    2.536
## 7256 84864614709 4.765 2012    1700     2    3.324
## 7611 84856575269 3.741 2012    2600     1    2.595
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55672 -0.46097  0.00656  0.46908  3.32363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0382    0.0560   18.54 < 2e-16 ***
## FirstAuthorFemale1 -0.1701    0.0497   -3.42  0.00063 ***
## LastAuthorFemale1 -0.0600    0.0542   -1.11  0.26806
## UniqueAuthors2     0.1752    0.0317    5.52 3.6e-08 ***
## UniqueAuthors3     0.2957    0.0492    6.01 2.1e-09 ***
## UniqueAuthors4     0.5059    0.0820    6.17 7.8e-10 ***
## UniqueAuthors5     0.5874    0.0720    8.16 4.9e-16 ***
## Year1997        -0.1248    0.1336   -0.93  0.35033
```

```

## Year1998          0.1996      0.1561      1.28  0.20111
## Year1999          0.1484      0.0785      1.89  0.05878 .
## Year2000          0.0632      0.1855      0.34  0.73341
## Year2001          0.0536      0.0833      0.64  0.52007
## Year2002          0.1869      0.0883      2.12  0.03436 *
## Year2003          0.1132      0.0739      1.53  0.12587
## Year2004          0.0548      0.0734      0.75  0.45544
## Year2005          0.0310      0.0757      0.41  0.68228
## Year2006          0.0336      0.0693      0.48  0.62825
## Year2007          0.0796      0.0684      1.16  0.24415
## Year2008         -0.0264      0.0689     -0.38  0.70175
## Year2009         -0.0689      0.0722     -0.95  0.34024
## Year2010          0.0926      0.0714      1.30  0.19522
## Year2011         -0.0206      0.0734     -0.28  0.77891
## Year2012          0.1075      0.0776      1.39  0.16581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.0566, Adjusted R-squared:  0.0496
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2779 is an outlier with |weight| = 0 ( < 3.4e-05);
## 269 weights are ~= 1. The remaining 2715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.119  0.866  0.949   0.910   0.986   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.35e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.788 1         1.337
## LastAuthorFemale  1.791 1         1.338
## Year              1.063 16         1.002

```

## Residuals from first and last author



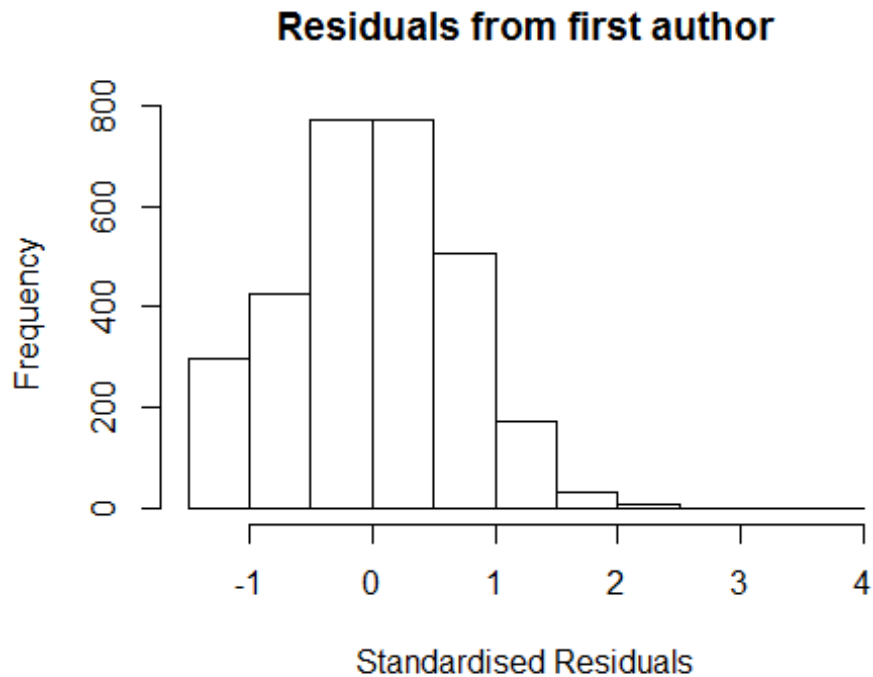
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7256 84864614709 4.765 2012    1700    2    3.511
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29258 -0.47185 -0.00253  0.48215  3.51115
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08753    0.05683   19.14  <2e-16 ***
## FirstAuthorFemale1 -0.13018    0.05172   -2.52   0.012 *
## LastAuthorFemale1 -0.07629    0.05716   -1.33   0.182
## Year1997        -0.13266    0.13040   -1.02   0.309
## Year1998         0.20720    0.15429    1.34   0.179
## Year1999         0.18334    0.07997    2.29   0.022 *
## Year2000         0.05165    0.21109    0.24   0.807
## Year2001         0.05484    0.08487    0.65   0.518
## Year2002         0.20505    0.09012    2.28   0.023 *
## Year2003         0.13532    0.07557    1.79   0.073 .
## Year2004         0.08000    0.07497    1.07   0.286
## Year2005         0.06913    0.07699    0.90   0.369
```

```

## Year2006          0.06503      0.07042      0.92      0.356
## Year2007          0.11008      0.06924      1.59      0.112
## Year2008         -0.00444      0.07005     -0.06      0.950
## Year2009         -0.00168      0.07356     -0.02      0.982
## Year2010          0.14131      0.07173      1.97      0.049 *
## Year2011          0.02472      0.07449      0.33      0.740
## Year2012          0.16631      0.07775      2.14      0.033 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.0124
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2779 is an outlier with |weight| = 0 ( < 3.4e-05);
## 275 weights are ~= 1. The remaining 2709 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.181  0.864   0.950   0.911   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## Year              1.037 16          1.001

```



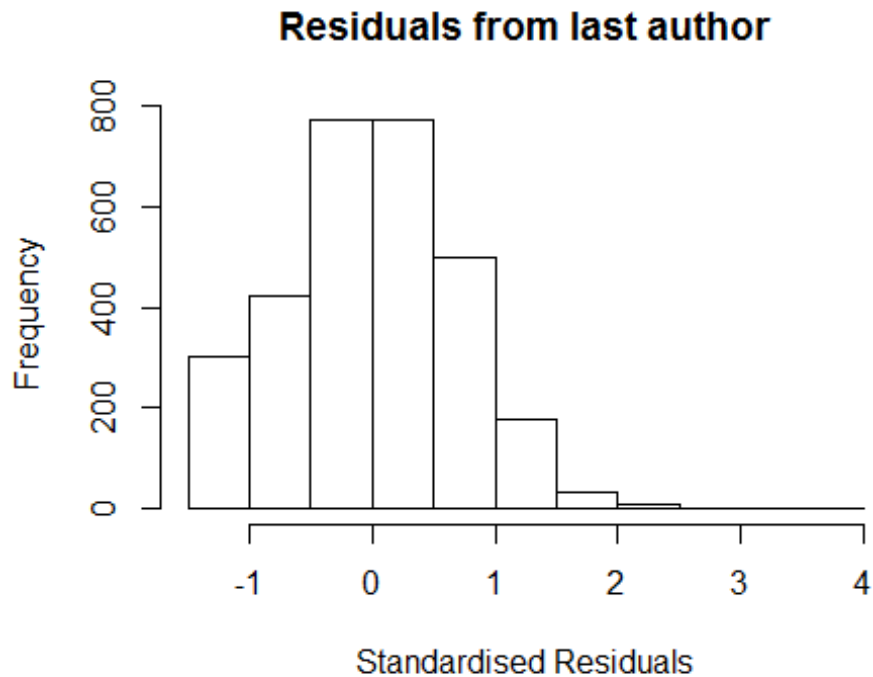


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7256 84864614709 4.765 2012    1700      2    3.511
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.289635 -0.472299 -0.000401  0.484166  3.512995
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08508    0.05659   19.17 < 2e-16 ***
## FirstAuthorFemale1 -0.17397    0.03983   -4.37 1.3e-05 ***
## Year1997      -0.12809    0.12974   -0.99  0.324
## Year1998       0.20302    0.15264    1.33  0.184
## Year1999       0.18100    0.07985    2.27  0.023 *
## Year2000       0.04057    0.21211    0.19  0.848
## Year2001       0.05224    0.08458    0.62  0.537
## Year2002       0.20455    0.09020    2.27  0.023 *
## Year2003       0.13422    0.07535    1.78  0.075 .
## Year2004       0.07922    0.07478    1.06  0.289
## Year2005       0.06711    0.07680    0.87  0.382
## Year2006       0.06604    0.07034    0.94  0.348
```

```

## Year2007          0.11032      0.06913      1.60      0.111
## Year2008          -0.00340      0.06996     -0.05      0.961
## Year2009          -0.00325      0.07336     -0.04      0.965
## Year2010           0.14175      0.07154      1.98      0.048 *
## Year2011           0.02473      0.07444      0.33      0.740
## Year2012           0.16693      0.07760      2.15      0.032 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.703
## Multiple R-squared:  0.0176, Adjusted R-squared:  0.012
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2779 is an outlier with |weight| = 0 ( < 3.4e-05);
## 265 weights are ~= 1. The remaining 2719 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.184  0.865  0.950  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.04 1          1.020
## Year              1.04 16          1.001

```



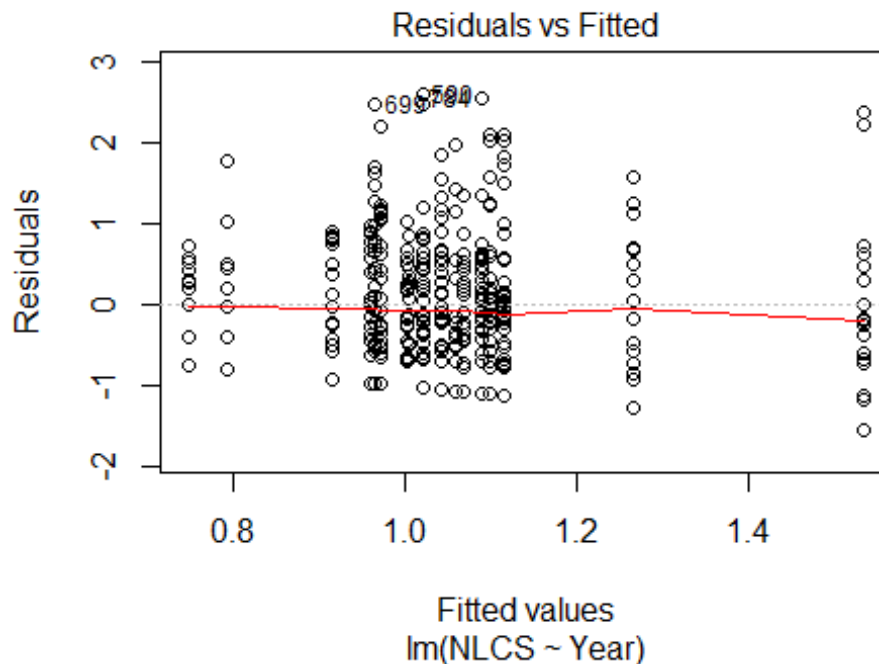
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 7256 84864614709 4.765 2012    1700      2    3.511
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.28573 -0.47012  0.00345  0.48524  3.52088
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08523    0.05682   19.10 < 2e-16 ***
## LastAuthorFemale1 -0.16137    0.04452   -3.62  0.00029 ***
## Year1997       -0.13647    0.13037   -1.05  0.29529
## Year1998        0.21105    0.15697    1.34  0.17889
## Year1999        0.18168    0.08005    2.27  0.02331 *
## Year2000        0.03530    0.21302    0.17  0.86840
## Year2001        0.05247    0.08520    0.62  0.53805
## Year2002        0.20050    0.08991    2.23  0.02582 *
## Year2003        0.13513    0.07591    1.78  0.07516 .
## Year2004        0.07326    0.07526    0.97  0.33046
## Year2005        0.06893    0.07700    0.90  0.37078
## Year2006        0.05304    0.07016    0.76  0.44972
```

```

## Year2007      0.10632      0.06925      1.54  0.12482
## Year2008     -0.00974      0.07001     -0.14  0.88936
## Year2009     -0.00774      0.07352     -0.11  0.91616
## Year2010      0.13427      0.07177      1.87  0.06147 .
## Year2011      0.01579      0.07417      0.21  0.83142
## Year2012      0.15889      0.07769      2.05  0.04092 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.016, Adjusted R-squared:  0.0104
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2779 is an outlier with |weight| = 0 ( < 3.4e-05);
## 269 weights are ~= 1. The remaining 2715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.177  0.864  0.950  0.911  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2985"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2601"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   33   41   20   32   43   23   32   32   38   51   52   69   51   53
## 2011 2012
##   54   54
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   15   16   11   22   21   16   14   19   17   35   32   38   40   38

```

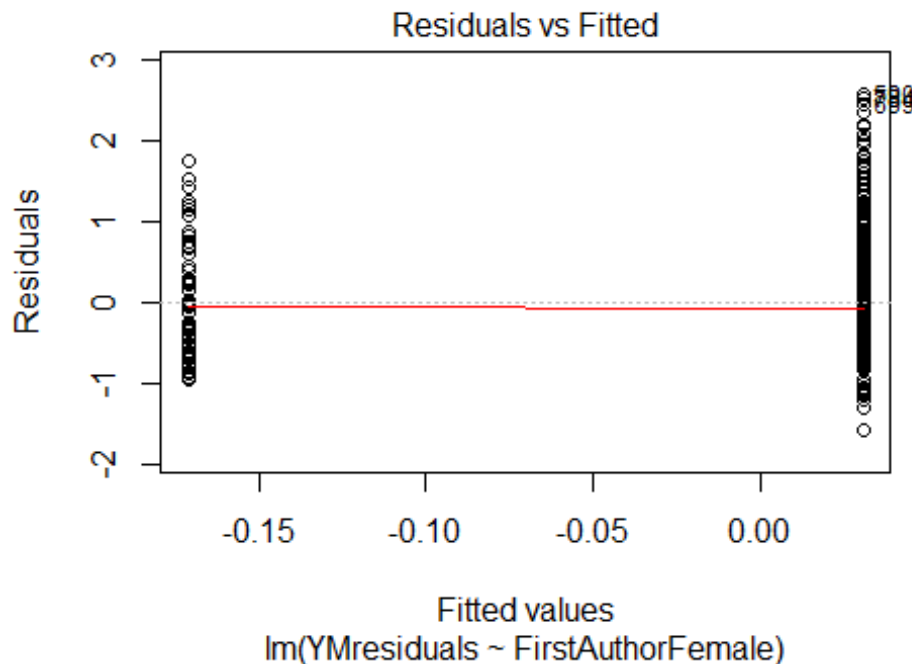
```
## 2011 2012
## 35 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 14 16 11 22 20 13 13 17 16 35 29 35 39 36
## 2011 2012
## 34 37
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.09
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 1.5157165665104"
## [1] "Male first author team size 2018 geometric mean: 1.23056125764876"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 72, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.25992104989487"
## [1] "Male last author team size 2018 geometric mean: 1.28138745447418"

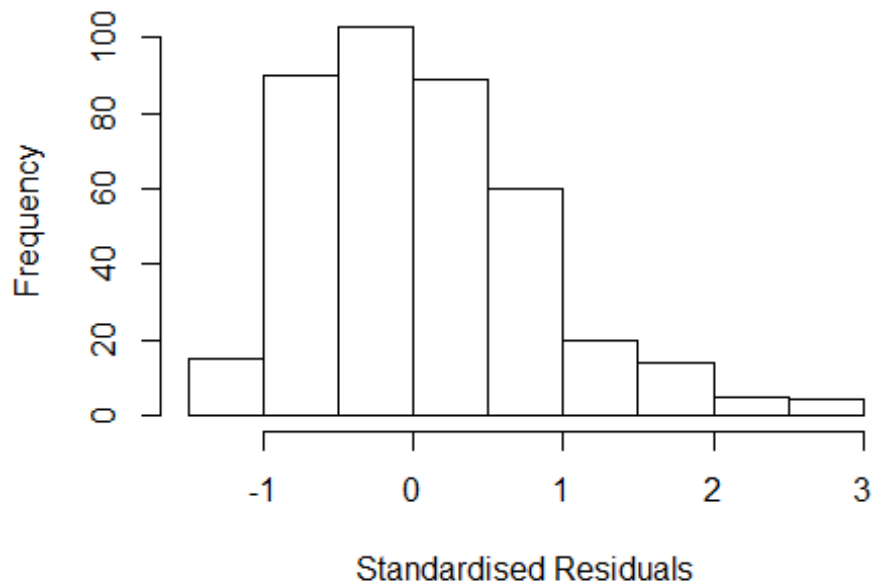
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 36, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.872	1	1.368
LastAuthorFemale	1.868	1	1.367
UniqueAuthors	1.994	4	1.090
Year	2.454	16	1.028

## Residuals from first and last author and team size



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 354 33646501982 3.908 2005    2601      1    2.684
## 374 32444450470 3.757 2005    2601      1    2.533
## 699 77957671687 3.447 2010    2601      1    2.673
## 784 79960824714 3.652 2011    2601      1    2.739
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.224 -0.519 -0.032  0.519  2.739
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6639    0.2275   2.92  0.0037 **
## FirstAuthorFemale1 -0.3283    0.1327  -2.48  0.0138 *
## LastAuthorFemale1  0.1149    0.1375   0.84  0.4039
## UniqueAuthors2     0.2905    0.1007   2.88  0.0041 **
## UniqueAuthors3     0.5246    0.1247   4.21 3.2e-05 ***
## UniqueAuthors4     0.5886    0.3455   1.70  0.0893 .
## UniqueAuthors5     1.0837    0.6763   1.60  0.1099
```

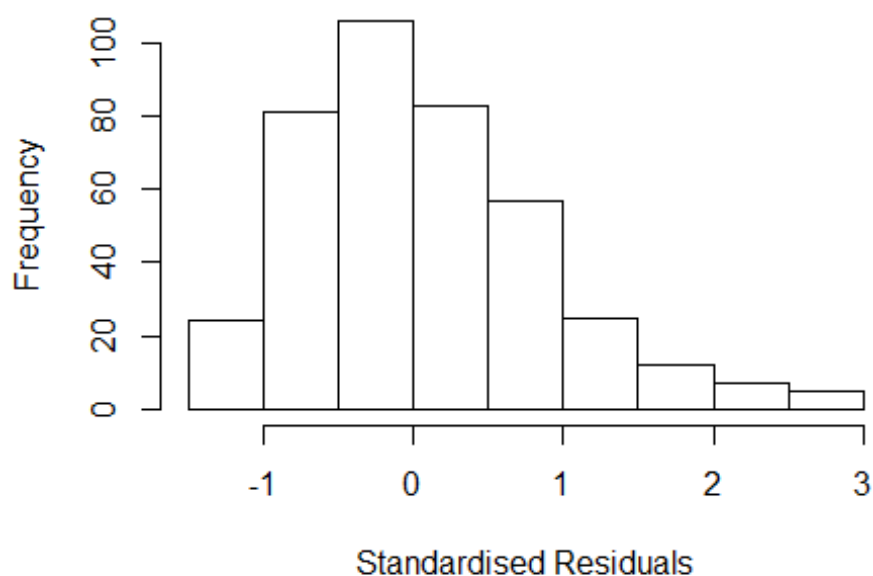
```

## Year1997          0.0542      0.3150      0.17      0.8634
## Year1998          0.3728      0.3515      1.06      0.2895
## Year1999          0.2375      0.2871      0.83      0.4087
## Year2000          0.3014      0.2629      1.15      0.2523
## Year2001          0.1139      0.2619      0.44      0.6638
## Year2002          0.0352      0.2671      0.13      0.8953
## Year2003          0.1673      0.2741      0.61      0.5418
## Year2004          0.1405      0.2985      0.47      0.6382
## Year2005          0.5599      0.3077      1.82      0.0697 .
## Year2006          0.0705      0.2783      0.25      0.8003
## Year2007          0.2882      0.2786      1.03      0.3017
## Year2008          0.1978      0.2693      0.73      0.4630
## Year2009          0.2381      0.2549      0.93      0.3508
## Year2010          0.1105      0.2735      0.40      0.6864
## Year2011          0.2492      0.2539      0.98      0.3271
## Year2012          0.2698      0.2577      1.05      0.2957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.741
## Multiple R-squared:  0.0969, Adjusted R-squared:  0.0442
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 373 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.143  0.870  0.948  0.901  0.984  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.50e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.724 1           1.313
## LastAuthorFemale 1.691 1           1.300
## Year              1.324 16           1.009

```



## Residuals from first and last author



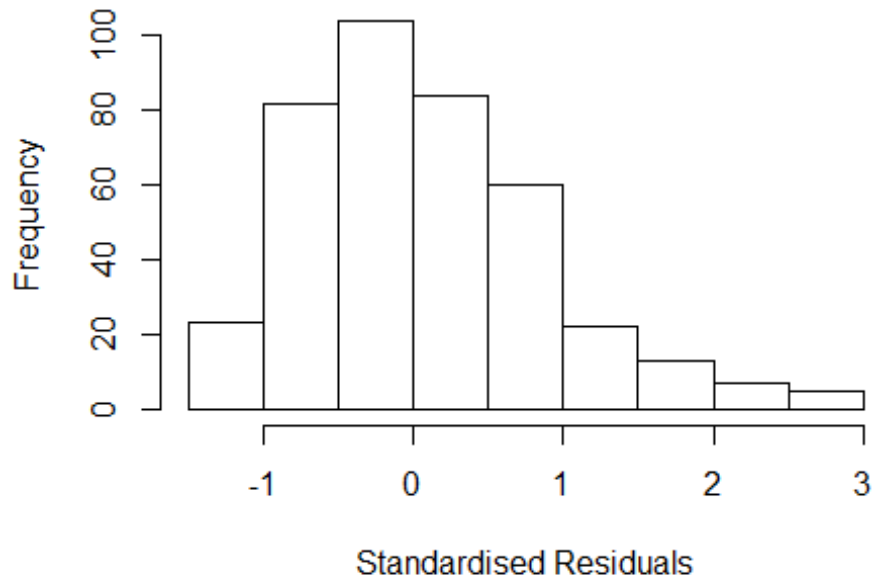
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 354 33646501982 3.908 2005    2601      1    2.541
## 531 43749120910 3.501 2008    2601      1    2.557
## 590 43749097000 3.620 2008    2601      1    2.676
## 699 77957671687 3.447 2010    2601      1    2.543
## 784 79960824714 3.652 2011    2601      1    2.630
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3670 -0.5138 -0.0477  0.5308  2.6757
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7750     0.2017   3.84 0.00014 ***
## FirstAuthorFemale1 -0.2669     0.1331  -2.00 0.04572 *
## LastAuthorFemale1  0.0800     0.1399   0.57 0.56784
## Year1997         0.0971     0.3078   0.32 0.75251
## Year1998         0.4936     0.3280   1.50 0.13317
## Year1999         0.2275     0.2486   0.92 0.36072
## Year2000         0.3149     0.2411   1.31 0.19230
## Year2001         0.1885     0.2542   0.74 0.45884
```

```

## Year2002          0.1506      0.2490      0.60  0.54568
## Year2003          0.0748      0.2560      0.29  0.77036
## Year2004          0.1066      0.2713      0.39  0.69469
## Year2005          0.5920      0.3156      1.88  0.06144 .
## Year2006          0.1607      0.2584      0.62  0.53448
## Year2007          0.2477      0.2623      0.94  0.34553
## Year2008          0.1693      0.2520      0.67  0.50207
## Year2009          0.2801      0.2370      1.18  0.23797
## Year2010          0.1287      0.2496      0.52  0.60636
## Year2011          0.2467      0.2351      1.05  0.29468
## Year2012          0.2840      0.2410      1.18  0.23930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.772
## Multiple R-squared:  0.0337, Adjusted R-squared:  -0.0119
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 371 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.205  0.876  0.949  0.902  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.113 1      1.055
## Year              1.113 16      1.003

```

## Residuals from first author



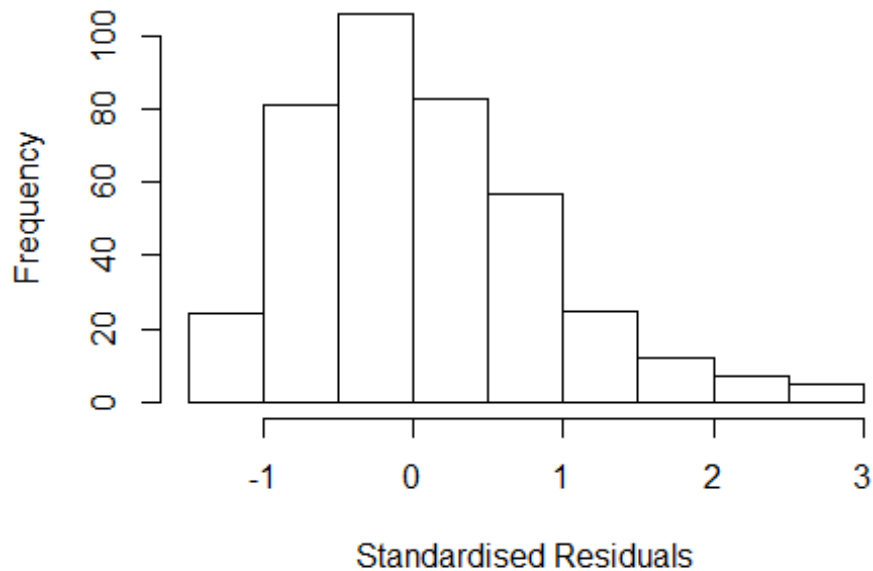
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 354 33646501982 3.908 2005    2601      1    2.541
## 531 43749120910 3.501 2008    2601      1    2.557
## 590 43749097000 3.620 2008    2601      1    2.676
## 699 77957671687 3.447 2010    2601      1    2.543
## 784 79960824714 3.652 2011    2601      1    2.630
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3798 -0.5288 -0.0434  0.5402  2.6680
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7666    0.2032   3.77 0.00019 ***
## FirstAuthorFemale1 -0.2200    0.1069  -2.06 0.04026 *
## Year1997        0.1130    0.3065   0.37 0.71253
## Year1998        0.5128    0.3288   1.56 0.11963
## Year1999        0.2359    0.2499   0.94 0.34575
## Year2000        0.3370    0.2398   1.41 0.16081
## Year2001        0.1984    0.2548   0.78 0.43657
## Year2002        0.1618    0.2471   0.66 0.51282
```

```

## Year2003          0.0886      0.2571      0.34  0.73074
## Year2004          0.1200      0.2695      0.45  0.65627
## Year2005          0.6131      0.3102      1.98  0.04881 *
## Year2006          0.1712      0.2597      0.66  0.51004
## Year2007          0.2542      0.2632      0.97  0.33481
## Year2008          0.1854      0.2501      0.74  0.45902
## Year2009          0.2842      0.2395      1.19  0.23620
## Year2010          0.1378      0.2511      0.55  0.58342
## Year2011          0.2612      0.2351      1.11  0.26719
## Year2012          0.2966      0.2427      1.22  0.22252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.77
## Multiple R-squared:  0.0331, Adjusted R-squared:  -0.00994
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~ = 1. The remaining 372 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.206  0.878   0.950   0.902   0.982   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           2.50e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##           500           50         2         1         1000         200
## trace.lev      mts    compute.rd
##           0           1000         0
##           psi                subsampling                cov
##           "bisquare"          "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.095 1         1.046
## Year              1.095 16         1.003

```

## Residuals from last author



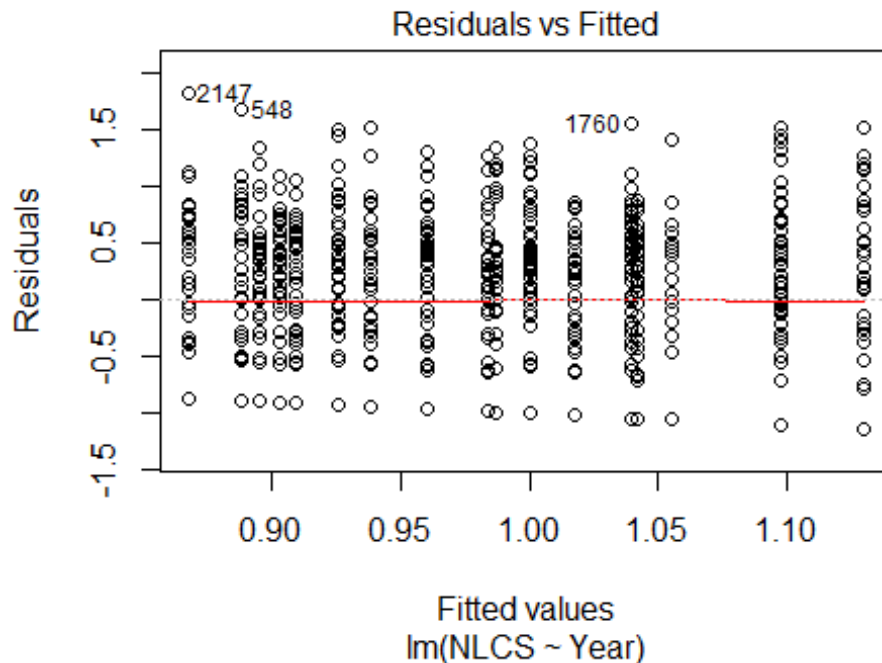
```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 354 33646501982 3.908 2005    2601      1    2.541
## 531 43749120910 3.501 2008    2601      1    2.557
## 590 43749097000 3.620 2008    2601      1    2.676
## 699 77957671687 3.447 2010    2601      1    2.543
## 784 79960824714 3.652 2011    2601      1    2.630
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3757 -0.5310 -0.0374  0.5764  2.6746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7333     0.2153   3.41 0.00073 ***
## LastAuthorFemale1 -0.0941     0.1153  -0.82 0.41475
## Year1997         0.1617     0.3168   0.51 0.61015
## Year1998         0.5316     0.3399   1.56 0.11858
## Year1999         0.2693     0.2597   1.04 0.30031
## Year2000         0.3546     0.2504   1.42 0.15748
## Year2001         0.1932     0.2645   0.73 0.46563
## Year2002         0.1884     0.2516   0.75 0.45442
```

```

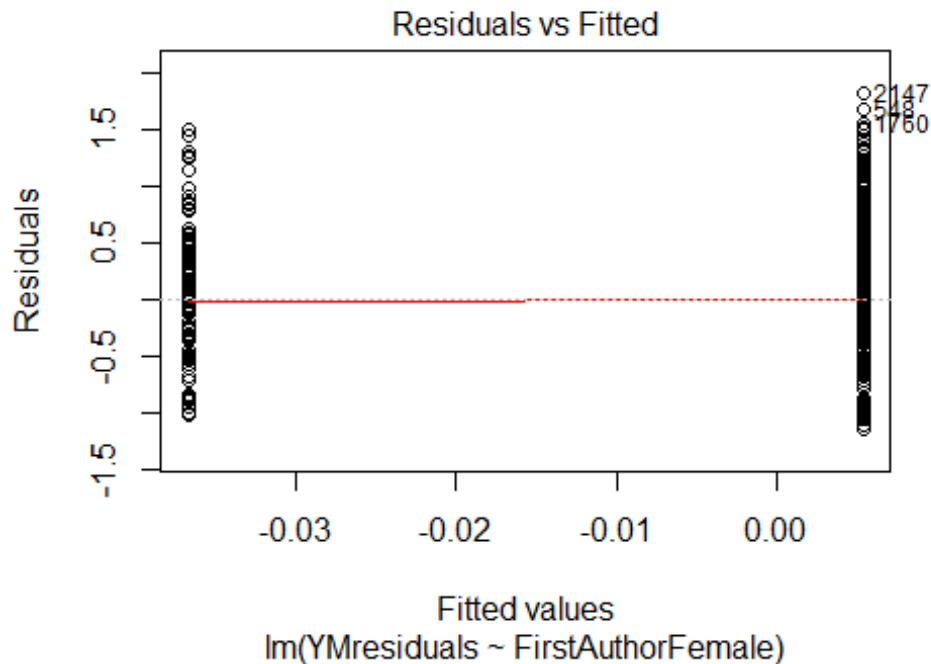
## Year2003      0.1008      0.2656      0.38  0.70450
## Year2004      0.1203      0.2782      0.43  0.66559
## Year2005      0.6424      0.3277      1.96  0.05069 .
## Year2006      0.1860      0.2714      0.69  0.49352
## Year2007      0.2641      0.2713      0.97  0.33079
## Year2008      0.2121      0.2596      0.82  0.41443
## Year2009      0.2712      0.2525      1.07  0.28329
## Year2010      0.1542      0.2644      0.58  0.56015
## Year2011      0.2847      0.2476      1.15  0.25089
## Year2012      0.3023      0.2547      1.19  0.23596
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.784
## Multiple R-squared:  0.0248, Adjusted R-squared:  -0.0186
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~ = 1. The remaining 369 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.220  0.879   0.947   0.904   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.50e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 400"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2602"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   85   71   80   92   97   92   94   85   83   76  111  103  115  128  124
## 2011 2012
##  157  129

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   45   33   49   61   69   55   60   46   60   55   68   72   72   88   85
## 2011 2012
##  112   94
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   33   46   57   62   53   56   42   59   54   65   70   69   83   84
## 2011 2012
##  106   88
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



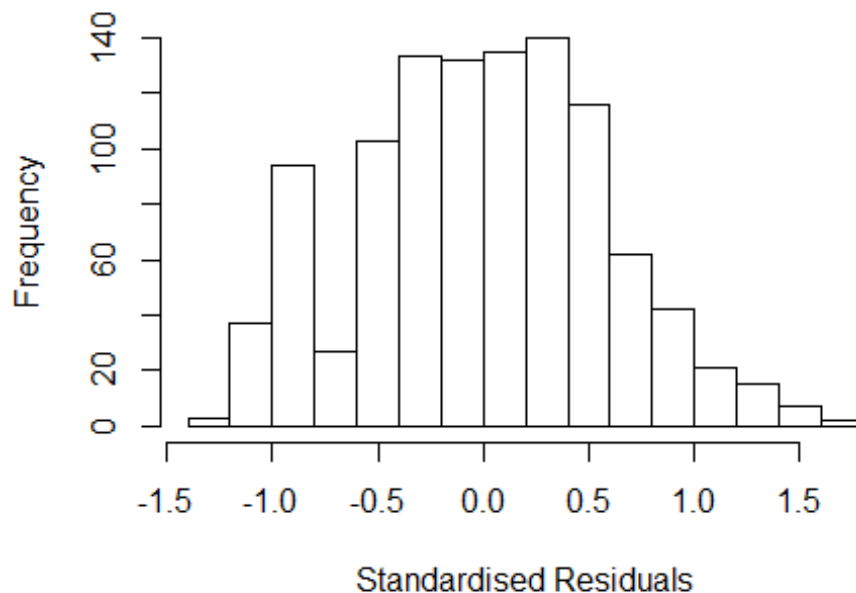
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.019, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 1.70445399394012"
## [1] "Male first author team size 2018 geometric mean: 1.22460999157944"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.51385904479011"
## [1] "Male last author team size 2018 geometric mean: 1.28518025735501"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 750, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.083  1      1.443
## LastAuthorFemale  2.086  1      1.444
## UniqueAuthors    1.227  3      1.035
## Year             1.299 16      1.008
```



## Residuals from first and last author and team size



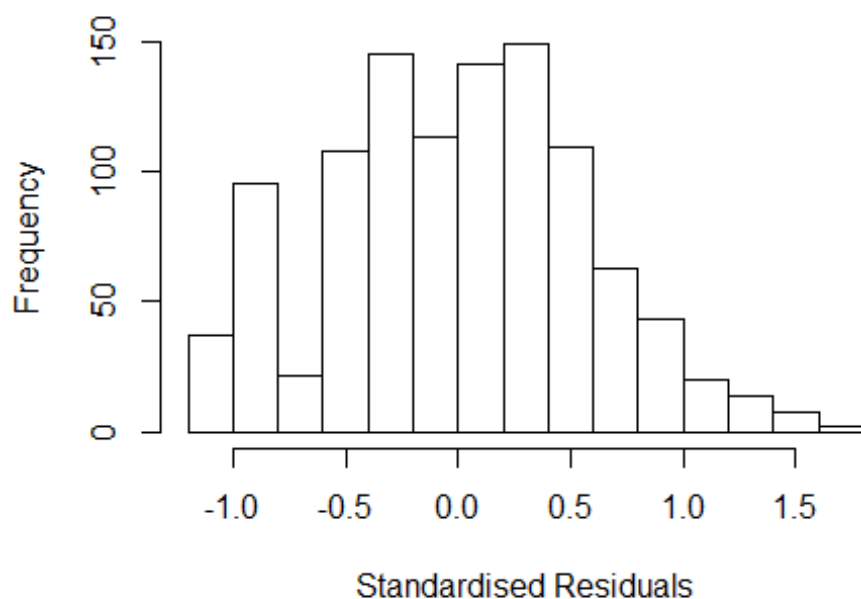
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3905 -0.3944  0.0113  0.3948  1.7448
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1607    0.1299   8.93  <2e-16 ***
## FirstAuthorFemale1 -0.1228    0.0777  -1.58   0.114
## LastAuthorFemale1  0.1105    0.0796   1.39   0.165
## UniqueAuthors2     0.0768    0.0469   1.64   0.101
## UniqueAuthors3     0.3123    0.1291   2.42   0.016 *
## UniqueAuthors4    -0.2064    0.2954  -0.70   0.485
## Year1997          -0.1231    0.1574  -0.78   0.435
## Year1998          -0.1596    0.1549  -1.03   0.303
## Year1999          -0.1930    0.1548  -1.25   0.213
## Year2000          -0.1212    0.1584  -0.76   0.445
```

```

## Year2001      -0.3365      0.1575      -2.14      0.033 *
## Year2002      -0.2355      0.1535      -1.53      0.125
## Year2003      -0.3039      0.1601      -1.90      0.058 .
## Year2004      -0.1203      0.1477      -0.81      0.415
## Year2005      -0.2614      0.1526      -1.71      0.087 .
## Year2006      -0.2846      0.1454      -1.96      0.051 .
## Year2007      -0.2705      0.1441      -1.88      0.061 .
## Year2008      -0.2055      0.1430      -1.44      0.151
## Year2009      -0.2607      0.1446      -1.80      0.072 .
## Year2010      -0.1406      0.1490      -0.94      0.346
## Year2011      -0.1978      0.1432      -1.38      0.167
## Year2012      -0.2976      0.1466      -2.03      0.043 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0268, Adjusted R-squared:  0.00726
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 979 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.351  0.862  0.949  0.910  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.958 1      1.399
## LastAuthorFemale  1.960 1      1.400
## Year              1.083 16      1.002

```

## Residuals from first and last author



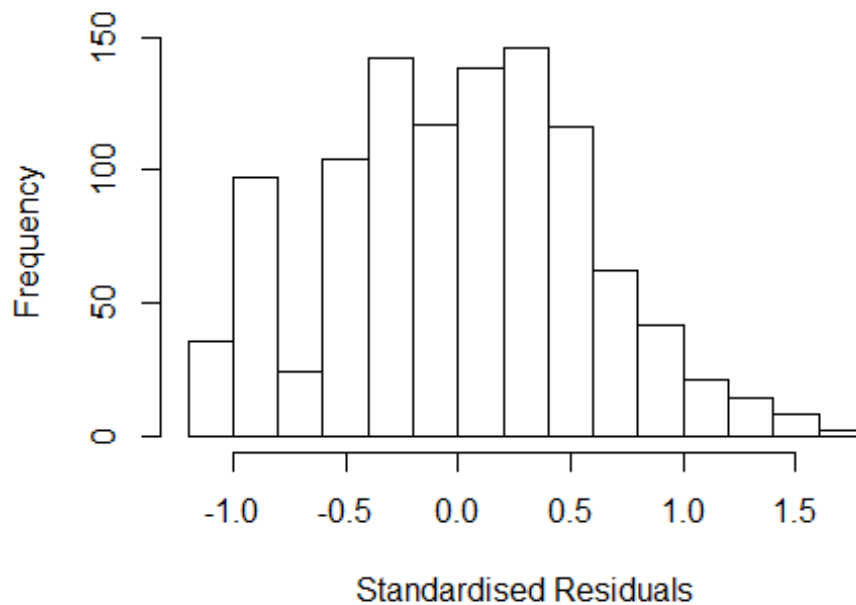
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1731 -0.3797  0.0261  0.3891  1.7993
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1731     0.1291   9.09  <2e-16 ***
## FirstAuthorFemale1 -0.1144     0.0759  -1.51   0.132
## LastAuthorFemale1  0.1093     0.0778   1.41   0.160
## Year1997          -0.1261     0.1557  -0.81   0.418
## Year1998          -0.1568     0.1540  -1.02   0.309
## Year1999          -0.1972     0.1543  -1.28   0.201
## Year2000          -0.1115     0.1568  -0.71   0.477
## Year2001          -0.3336     0.1561  -2.14   0.033 *
## Year2002          -0.2350     0.1537  -1.53   0.127
## Year2003          -0.2711     0.1592  -1.70   0.089 .
## Year2004          -0.1138     0.1463  -0.78   0.437
## Year2005          -0.2661     0.1522  -1.75   0.081 .
```

```

## Year2006          -0.2737      0.1455   -1.88    0.060 .
## Year2007          -0.2581      0.1435   -1.80    0.072 .
## Year2008          -0.2042      0.1423   -1.43    0.152
## Year2009          -0.2582      0.1435   -1.80    0.072 .
## Year2010          -0.1238      0.1483   -0.83    0.404
## Year2011          -0.1772      0.1414   -1.25    0.210
## Year2012          -0.2884      0.1457   -1.98    0.048 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.00189
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 975 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.326  0.866  0.947  0.910  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.034 1      1.017
## Year              1.034 16      1.001

```

## Residuals from first author



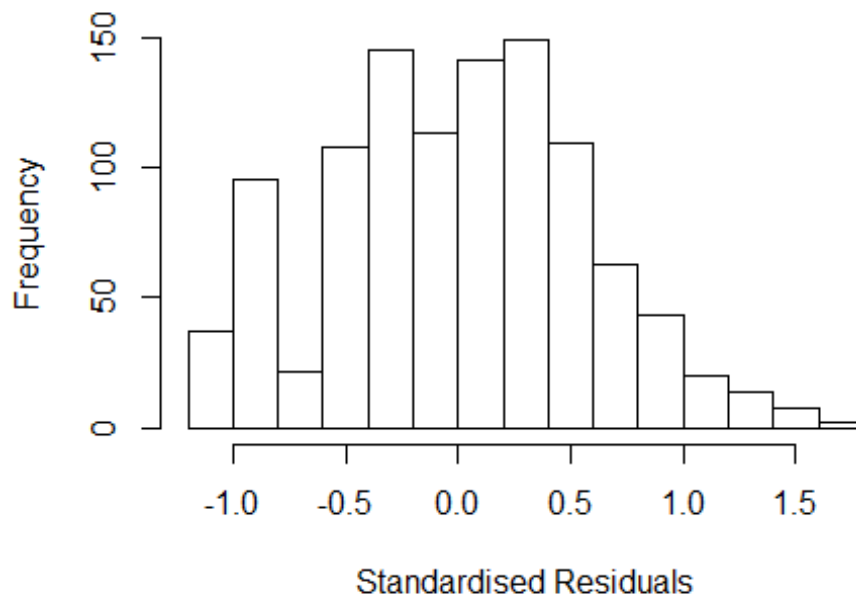
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1734 -0.3799  0.0224  0.3929  1.7929
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1734     0.1318   8.91  <2e-16 ***
## FirstAuthorFemale1 -0.0372     0.0553  -0.67   0.501
## Year1997         -0.1272     0.1574  -0.81   0.419
## Year1998         -0.1531     0.1567  -0.98   0.329
## Year1999         -0.1936     0.1564  -1.24   0.216
## Year2000         -0.1128     0.1589  -0.71   0.478
## Year2001         -0.3317     0.1578  -2.10   0.036 *
## Year2002         -0.2299     0.1563  -1.47   0.142
## Year2003         -0.2669     0.1623  -1.64   0.100
## Year2004         -0.1101     0.1489  -0.74   0.460
## Year2005         -0.2612     0.1545  -1.69   0.091 .
## Year2006         -0.2702     0.1478  -1.83   0.068 .
```

```

## Year2007          -0.2499      0.1459   -1.71    0.087 .
## Year2008          -0.2008      0.1447   -1.39    0.166
## Year2009          -0.2543      0.1459   -1.74    0.082 .
## Year2010          -0.1190      0.1504   -0.79    0.429
## Year2011          -0.1769      0.1439   -1.23    0.219
## Year2012          -0.2823      0.1482   -1.91    0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0169, Adjusted R-squared:  0.000955
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 971 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.330  0.862  0.946  0.909  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1          1.017
## Year            1.035 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1670 -0.3771  0.0238  0.3958  1.8019
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1670     0.1310   8.91  <2e-16 ***
## LastAuthorFemale1  0.0275     0.0570   0.48   0.629
## Year1997        -0.1248     0.1569  -0.80   0.426
## Year1998        -0.1566     0.1556  -1.01   0.314
## Year1999        -0.1941     0.1558  -1.25   0.213
## Year2000        -0.1142     0.1584  -0.72   0.471
## Year2001        -0.3308     0.1572  -2.10   0.036 *
## Year2002        -0.2317     0.1559  -1.49   0.138
## Year2003        -0.2757     0.1615  -1.71   0.088 .
## Year2004        -0.1131     0.1483  -0.76   0.446
## Year2005        -0.2611     0.1542  -1.69   0.091 .
## Year2006        -0.2718     0.1474  -1.84   0.065 .
```

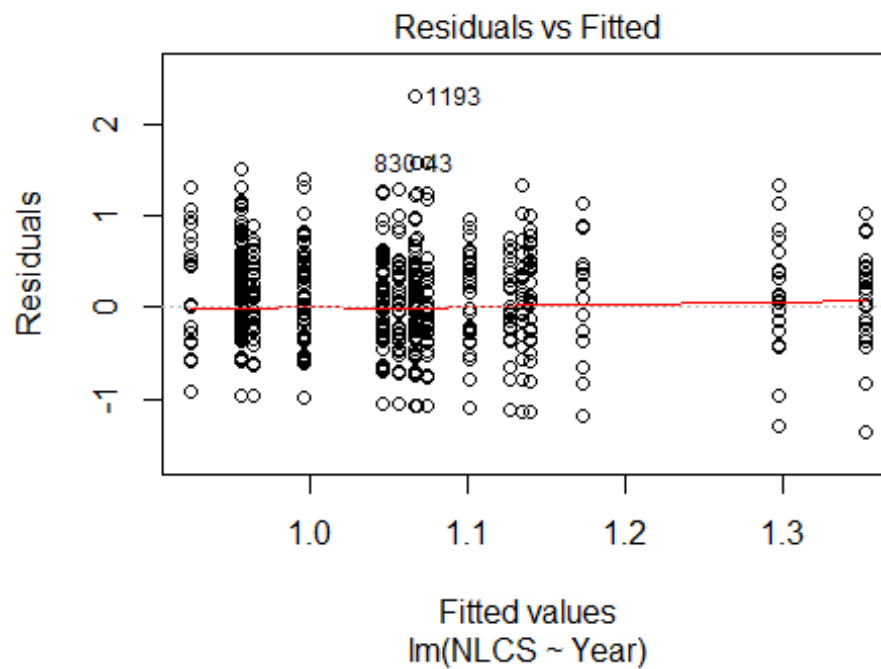
```

## Year2007          -0.2517      0.1453   -1.73    0.084 .
## Year2008          -0.2048      0.1441   -1.42    0.155
## Year2009          -0.2549      0.1456   -1.75    0.080 .
## Year2010          -0.1185      0.1496   -0.79    0.429
## Year2011          -0.1784      0.1432   -1.25    0.213
## Year2012          -0.2849      0.1476   -1.93    0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.587
## Multiple R-squared:  0.0166, Adjusted R-squared:  0.000719
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 973 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.327  0.862  0.948  0.910  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.35e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1069"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2603"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   81   62   64   61   96   68   64   85   70   81   87  114  131   99
## 2011 2012
##  109  114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   16   28   17   27   21   32   28   21   43   32   36   46   70   72   57
## 2011 2012

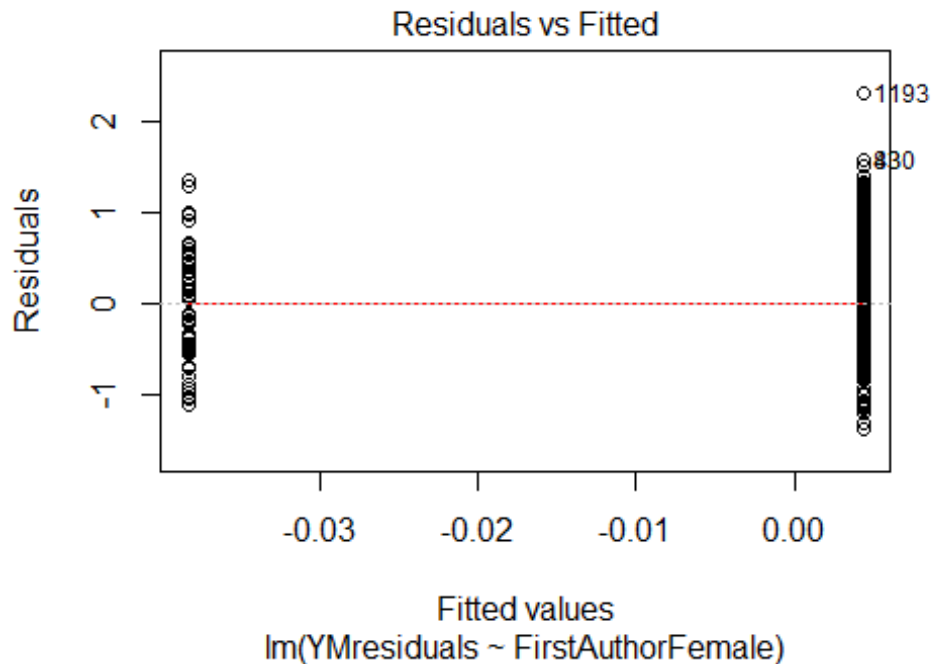
```



```
## 69 70
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 16 24 17 24 19 30 24 17 41 30 32 42 67 68 56
## 2011 2012
## 65 64
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```

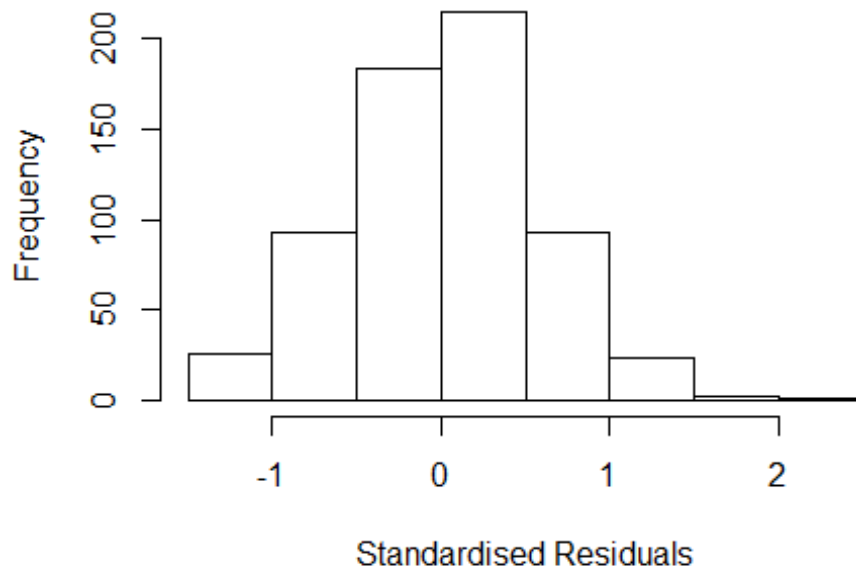


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.19, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 1.45242285611433"
## [1] "Male first author team size 2018 geometric mean: 1.48079594016579"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 370, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.71754006157419"
## [1] "Male last author team size 2018 geometric mean: 1.42153818312233"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 490, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.603 1          1.266
## LastAuthorFemale  1.648 1          1.284
## UniqueAuthors     1.790 4          1.076
## Year              2.090 16         1.023
```

## Residuals from first and last author and team size



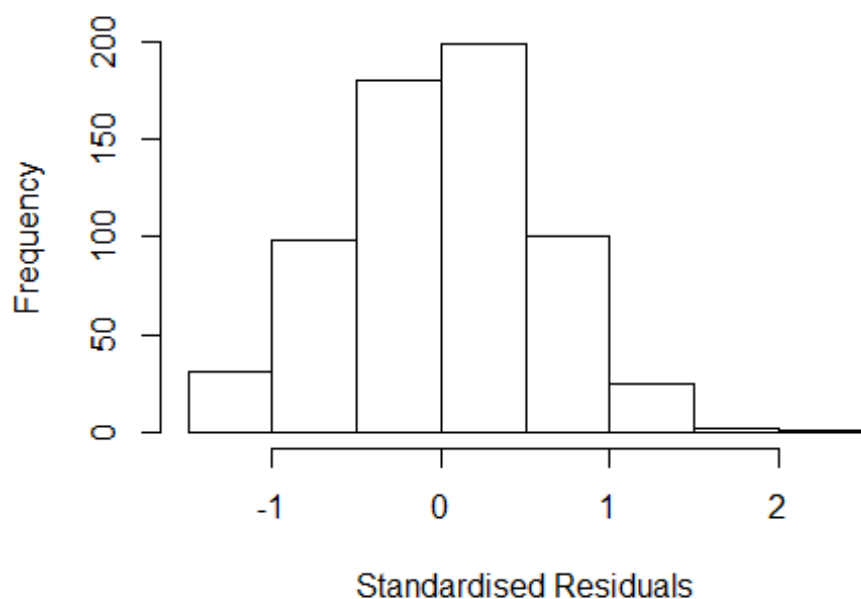
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3407 -0.3768 0.0291 0.3792 2.2573
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00268 0.17215 5.82 9.3e-09 ***
## FirstAuthorFemale1 0.00863 0.09590 0.09 0.9283
## LastAuthorFemale1 -0.10960 0.09926 -1.10 0.2700
## UniqueAuthors2 0.13196 0.05746 2.30 0.0220 *
## UniqueAuthors3 0.19002 0.11455 1.66 0.0977 .
## UniqueAuthors4 0.62027 0.24247 2.56 0.0108 *
## UniqueAuthors5 0.40148 0.12303 3.26 0.0012 **
## Year1997 0.14796 0.22380 0.66 0.5088
## Year1998 0.19938 0.26815 0.74 0.4574
## Year1999 0.31514 0.20898 1.51 0.1321
```

```

## Year2000      0.31085      0.25293      1.23      0.2195
## Year2001      0.02970      0.20550      0.14      0.8851
## Year2002     -0.19286      0.23737     -0.81      0.4168
## Year2003      0.08014      0.21524      0.37      0.7098
## Year2004      0.11277      0.19562      0.58      0.5645
## Year2005     -0.01115      0.19227     -0.06      0.9538
## Year2006     -0.14009      0.18947     -0.74      0.4599
## Year2007     -0.03367      0.19325     -0.17      0.8617
## Year2008     -0.02595      0.18413     -0.14      0.8880
## Year2009     -0.03804      0.18207     -0.21      0.8346
## Year2010     -0.10228      0.18380     -0.56      0.5781
## Year2011     -0.16342      0.18855     -0.87      0.3864
## Year2012     -0.07190      0.19129     -0.38      0.7072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0629, Adjusted R-squared:  0.0292
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 584 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0944 0.8690 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.576 1      1.255
## LastAuthorFemale 1.619 1      1.272
## Year      1.279 16      1.008

```

## Residuals from first and last author

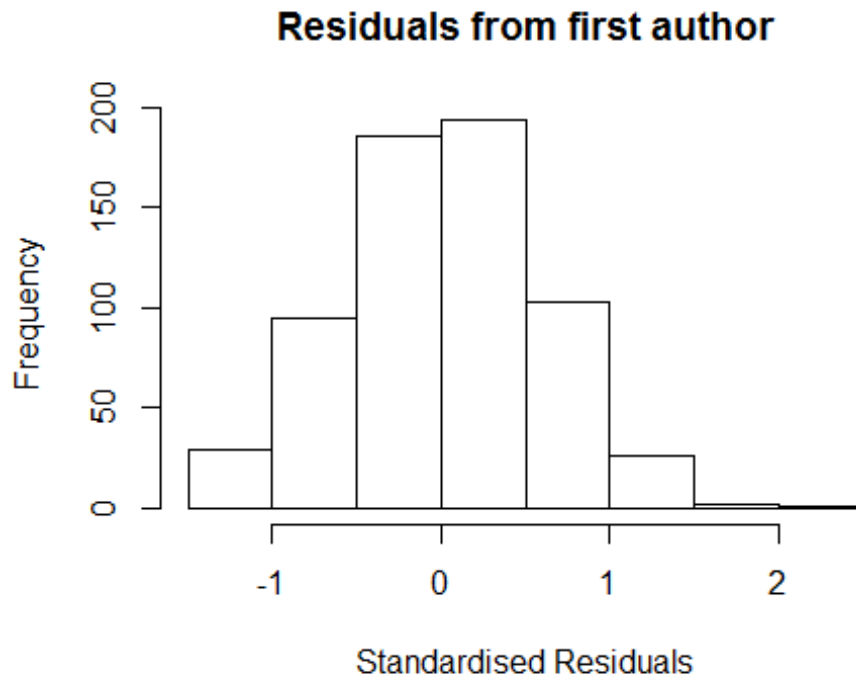


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3468 -0.3698 0.0204 0.3946 2.3460
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01374 0.17399 5.83 9.1e-09 ***
## FirstAuthorFemale1 0.03662 0.09760 0.38 0.71
## LastAuthorFemale1 -0.12320 0.10056 -1.23 0.22
## Year1997 0.17085 0.22297 0.77 0.44
## Year1998 0.20483 0.27910 0.73 0.46
## Year1999 0.32990 0.20917 1.58 0.12
## Year2000 0.33307 0.26153 1.27 0.20
## Year2001 0.06924 0.20779 0.33 0.74
## Year2002 -0.14370 0.24132 -0.60 0.55
## Year2003 0.10737 0.21177 0.51 0.61
## Year2004 0.14576 0.19846 0.73 0.46
## Year2005 0.00369 0.19498 0.02 0.98
```

```

## Year2006      -0.07972    0.19261   -0.41    0.68
## Year2007      0.01437    0.19286    0.07    0.94
## Year2008      0.00626    0.18660    0.03    0.97
## Year2009      0.01783    0.18363    0.10    0.92
## Year2010     -0.05764    0.18569   -0.31    0.76
## Year2011     -0.11242    0.19030   -0.59    0.55
## Year2012     -0.01496    0.19252   -0.08    0.94
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.0394, Adjusted R-squared:  0.0114
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 580 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0721 0.8770 0.9490 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.154 1      1.074
## Year      1.154 16      1.004

```



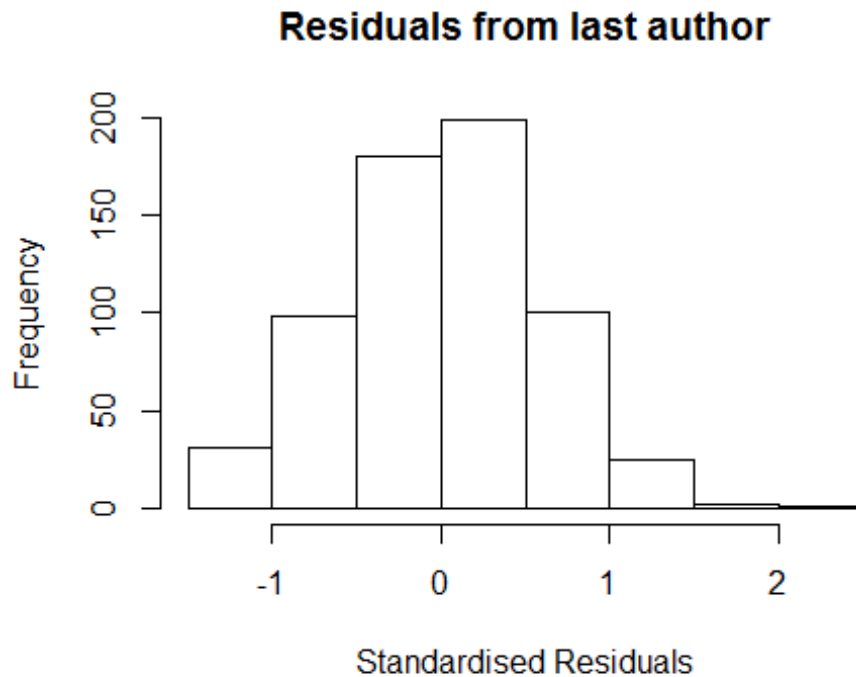
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3393 -0.3700 0.0235 0.4020 2.3517
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00424 0.17313 5.80 1.1e-08 ***
## FirstAuthorFemale1 -0.03522 0.08517 -0.41 0.68
## Year1997 0.17996 0.22272 0.81 0.42
## Year1998 0.20060 0.27891 0.72 0.47
## Year1999 0.33507 0.20920 1.60 0.11
## Year2000 0.32098 0.25573 1.26 0.21
## Year2001 0.07383 0.20714 0.36 0.72
## Year2002 -0.16243 0.23942 -0.68 0.50
## Year2003 0.11575 0.21036 0.55 0.58
## Year2004 0.15190 0.19773 0.77 0.44
## Year2005 0.00934 0.19552 0.05 0.96
## Year2006 -0.07376 0.19244 -0.38 0.70
```

```

## Year2007          0.02447    0.19162    0.13    0.90
## Year2008          0.01002    0.18672    0.05    0.96
## Year2009          0.02744    0.18285    0.15    0.88
## Year2010         -0.05092    0.18482   -0.28    0.78
## Year2011         -0.10613    0.19013   -0.56    0.58
## Year2012         -0.00666    0.19224   -0.03    0.97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.0376, Adjusted R-squared:  0.0111
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 583 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0702 0.8800 0.9480 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.191 1          1.091
## Year            1.191 16          1.005

```





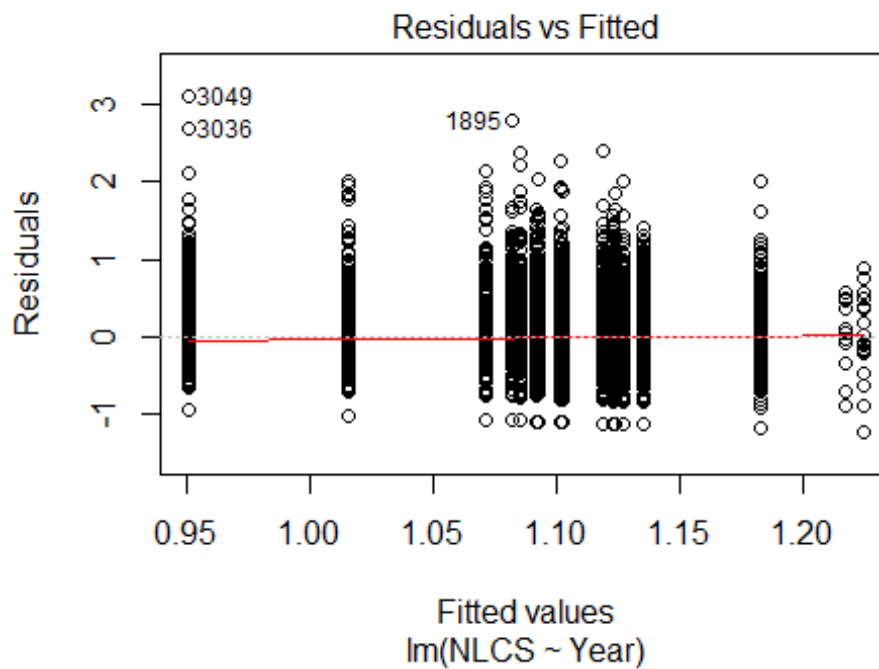
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3465 -0.3702 0.0193 0.3958 2.3454
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.01621 0.17356 5.86 7.7e-09 ***
## LastAuthorFemale1 -0.09963 0.08621 -1.16 0.25
## Year1997 0.16850 0.22236 0.76 0.45
## Year1998 0.20582 0.27904 0.74 0.46
## Year1999 0.32852 0.20892 1.57 0.12
## Year2000 0.33031 0.26100 1.27 0.21
## Year2001 0.06902 0.20803 0.33 0.74
## Year2002 -0.14720 0.24036 -0.61 0.54
## Year2003 0.10894 0.21247 0.51 0.61
## Year2004 0.14621 0.19875 0.74 0.46
## Year2005 0.00399 0.19536 0.02 0.98
## Year2006 -0.08123 0.19264 -0.42 0.67
```

```

## Year2007      0.01448    0.19275    0.08    0.94
## Year2008      0.00439    0.18636    0.02    0.98
## Year2009      0.01771    0.18370    0.10    0.92
## Year2010     -0.05953    0.18524   -0.32    0.75
## Year2011     -0.11303    0.19028   -0.59    0.55
## Year2012     -0.01696    0.19210   -0.09    0.93
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0394, Adjusted R-squared:  0.0129
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 584 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0726 0.8780 0.9490 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.57e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 636"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2604"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 765 709 693 43 71 840 757 590 609 584 667 785 790 849 789
## 2011 2012
## 745 718
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 239 232 210 11 19 277 321 224 265 254 323 397 394 434 406
## 2011 2012

```

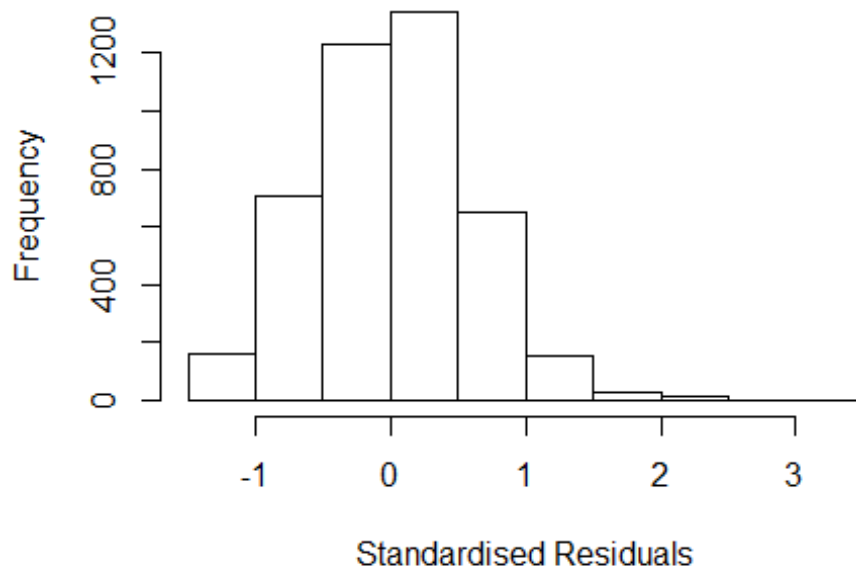
```
## 380 412
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 218 218 194 10 18 254 286 204 246 224 277 332 349 386 369
## 2011 2012
## 331 364
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 68, df = 16, p-value = 2e-08
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.9, df = 1, p-value = 0.05
```



## Residuals from first and last author and team size



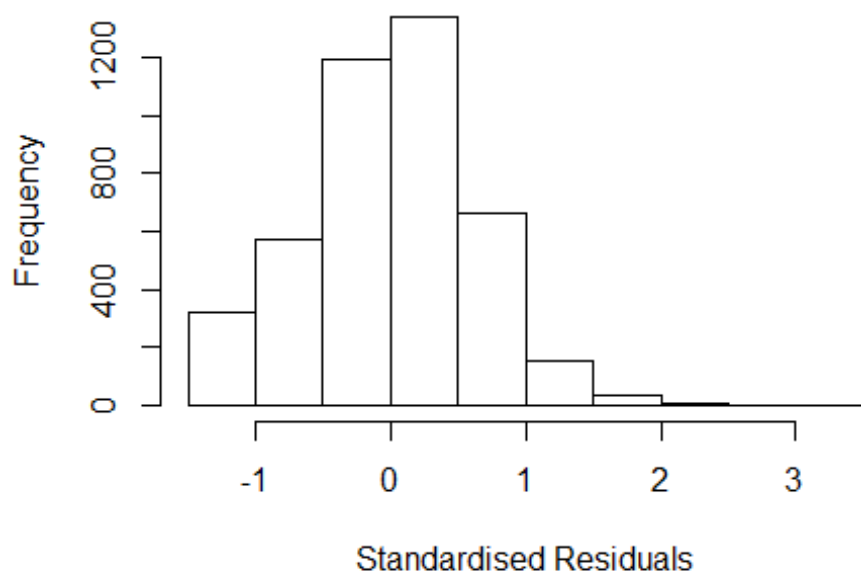
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1895 0001438979 3.878 1998    2002     2    2.647
## 3036 0035363218 3.647 2001    1702     5    2.634
## 3049 0035331869 4.055 2001    2604     1    3.303
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3403 -0.4010  0.0172  0.4007  3.3035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94482    0.04770   19.81 < 2e-16 ***
## FirstAuthorFemale1 -0.06425    0.02956   -2.17  0.0298 *
## LastAuthorFemale1 -0.02694    0.03191   -0.84  0.3986
## UniqueAuthors2    0.24809    0.02257   10.99 < 2e-16 ***
## UniqueAuthors3    0.26189    0.02854    9.18 < 2e-16 ***
## UniqueAuthors4    0.28083    0.04081    6.88 6.8e-12 ***
## UniqueAuthors5    0.29059    0.03608    8.05 1.0e-15 ***
## Year1997         0.01191    0.06541    0.18  0.8555
```

```

## Year1998      0.03844      0.06953      0.55      0.5804
## Year1999      0.23833      0.12838      1.86      0.0635 .
## Year2000      0.14795      0.14447      1.02      0.3059
## Year2001     -0.19331      0.06789     -2.85      0.0044 **
## Year2002     -0.10178      0.06236     -1.63      0.1027
## Year2003      0.10491      0.06220      1.69      0.0918 .
## Year2004      0.03642      0.06145      0.59      0.5534
## Year2005      0.02450      0.06255      0.39      0.6953
## Year2006      0.02214      0.05904      0.38      0.7076
## Year2007      0.04783      0.05727      0.84      0.4037
## Year2008     -0.02782      0.05596     -0.50      0.6192
## Year2009      0.00651      0.05588      0.12      0.9073
## Year2010      0.00883      0.05711      0.15      0.8772
## Year2011     -0.01921      0.05826     -0.33      0.7417
## Year2012     -0.02859      0.05760     -0.50      0.6197
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0534, Adjusted R-squared:  0.0485
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 750 is an outlier with |weight| = 0 ( < 2.3e-05);
## 347 weights are ~= 1. The remaining 3932 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0088 0.8670 0.9520 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.142 1      1.068
## LastAuthorFemale 1.138 1      1.067
## Year      1.027 16      1.001

```

## Residuals from first and last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1895 0001438979 3.878 1998    2002     2    2.803
## 3036 0035363218 3.647 2001    1702     5    2.743
## 3049 0035331869 4.055 2001    2604     1    3.151
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2608 -0.3995  0.0184  0.4091  3.1508
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0308    0.0492   20.93  <2e-16 ***
## FirstAuthorFemale1 -0.0281    0.0300   -0.94   0.349
## LastAuthorFemale1 -0.0251    0.0327   -0.77   0.443
## Year1997         0.0240    0.0681    0.35   0.725
## Year1998         0.0445    0.0725    0.61   0.539
## Year1999         0.2937    0.1333    2.20   0.028 *
## Year2000         0.2300    0.1422    1.62   0.106
## Year2001        -0.1266    0.0682   -1.86   0.063 .
## Year2002        -0.0368    0.0635   -0.58   0.562
## Year2003         0.1538    0.0634    2.42   0.015 *
```

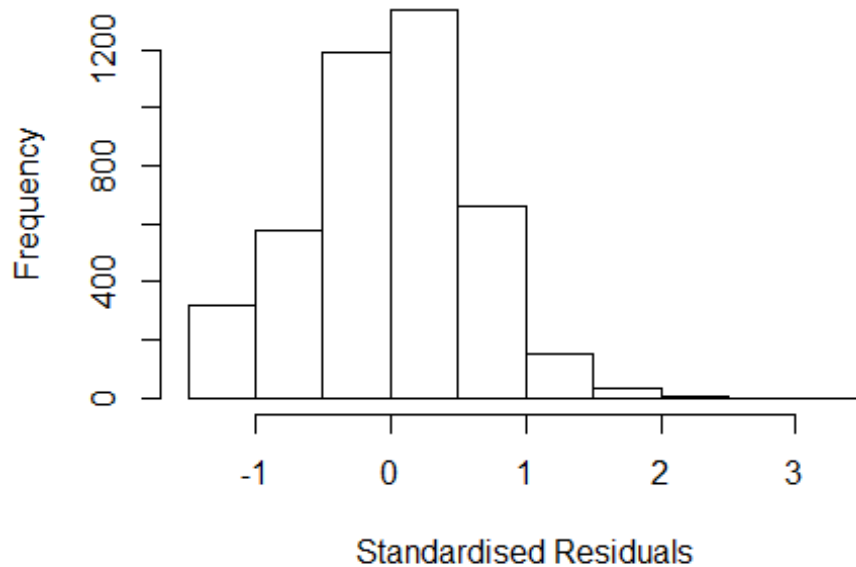
```

## Year2004          0.1020      0.0641      1.59      0.111
## Year2005          0.0916      0.0648      1.41      0.158
## Year2006          0.0711      0.0605      1.17      0.240
## Year2007          0.1034      0.0588      1.76      0.079 .
## Year2008          0.0406      0.0571      0.71      0.476
## Year2009          0.0713      0.0575      1.24      0.215
## Year2010          0.0838      0.0582      1.44      0.150
## Year2011          0.0562      0.0602      0.93      0.351
## Year2012          0.0499      0.0591      0.84      0.399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0109, Adjusted R-squared:  0.00672
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 750 is an outlier with |weight| = 0 ( < 2.3e-05);
## 356 weights are ~= 1. The remaining 3923 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0009 0.8700 0.9510 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.014 1          1.007
## Year              1.014 16          1.000

```



## Residuals from first author



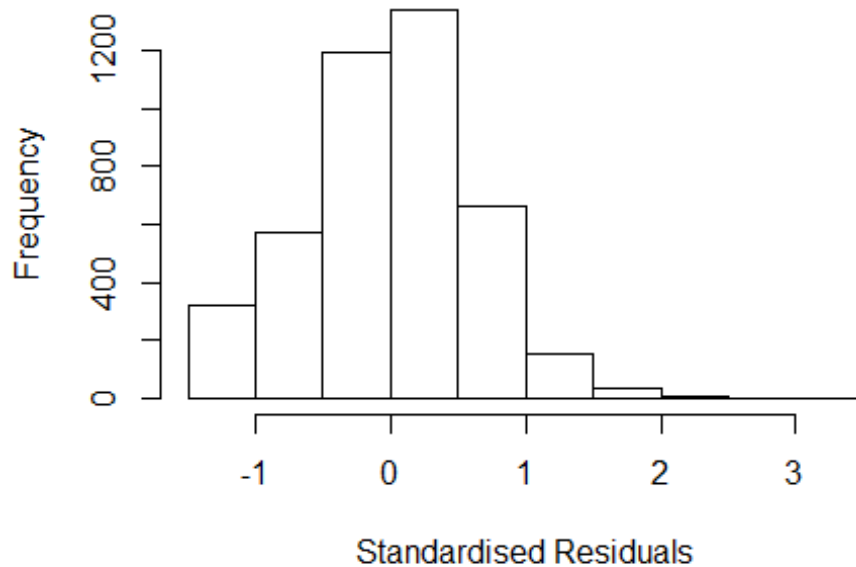
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1895 0001438979 3.878 1998    2002     2    2.803
## 3036 0035363218 3.647 2001    1702     5    2.743
## 3049 0035331869 4.055 2001    2604     1    3.151
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2617 -0.4004  0.0174  0.4096  3.1526
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0294    0.0492   20.94  <2e-16 ***
## FirstAuthorFemale1 -0.0362    0.0284   -1.27    0.203
## Year1997         0.0239    0.0681    0.35    0.725
## Year1998         0.0438    0.0724    0.61    0.545
## Year1999         0.2906    0.1322    2.20    0.028 *
## Year2000         0.2323    0.1429    1.63    0.104
## Year2001        -0.1270    0.0681   -1.86    0.062 .
## Year2002        -0.0373    0.0634   -0.59    0.557
## Year2003         0.1546    0.0634    2.44    0.015 *
## Year2004         0.1018    0.0640    1.59    0.112
```

```

## Year2005          0.0909      0.0648      1.40      0.160
## Year2006          0.0709      0.0605      1.17      0.241
## Year2007          0.1031      0.0588      1.75      0.080 .
## Year2008          0.0413      0.0570      0.72      0.469
## Year2009          0.0712      0.0574      1.24      0.215
## Year2010          0.0833      0.0582      1.43      0.152
## Year2011          0.0556      0.0602      0.92      0.356
## Year2012          0.0495      0.0591      0.84      0.402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00684
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 750 is an outlier with |weight| = 0 ( < 2.3e-05);
## 361 weights are ~= 1. The remaining 3918 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8700 0.9510 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.011 1      1.006
## Year      1.011 16      1.000

```

## Residuals from last author



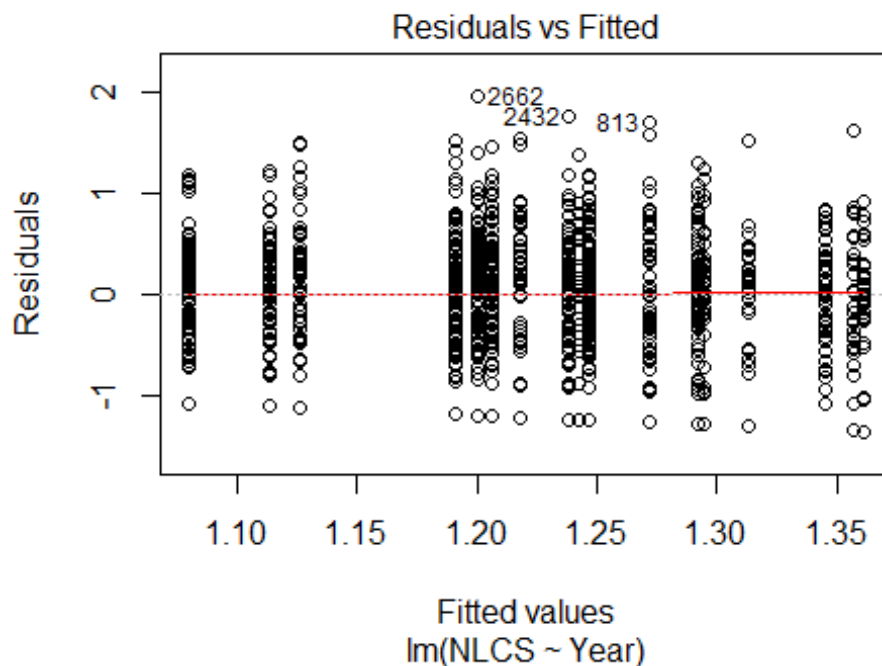
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1895 0001438979 3.878 1998    2002     2    2.803
## 3036 0035363218 3.647 2001    1702     5    2.743
## 3049 0035331869 4.055 2001    2604     1    3.151
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2551 -0.4001  0.0177  0.4086  3.1529
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0295     0.0493   20.88  <2e-16 ***
## LastAuthorFemale1 -0.0359     0.0310   -1.16    0.247
## Year1997          0.0234     0.0681    0.34    0.732
## Year1998          0.0434     0.0724    0.60    0.549
## Year1999          0.2942     0.1342    2.19    0.028 *
## Year2000          0.2256     0.1411    1.60    0.110
## Year2001         -0.1274     0.0682   -1.87    0.062 .
## Year2002         -0.0375     0.0635   -0.59    0.555
## Year2003          0.1515     0.0633    2.39    0.017 *
## Year2004          0.1008     0.0641    1.57    0.116
```

```

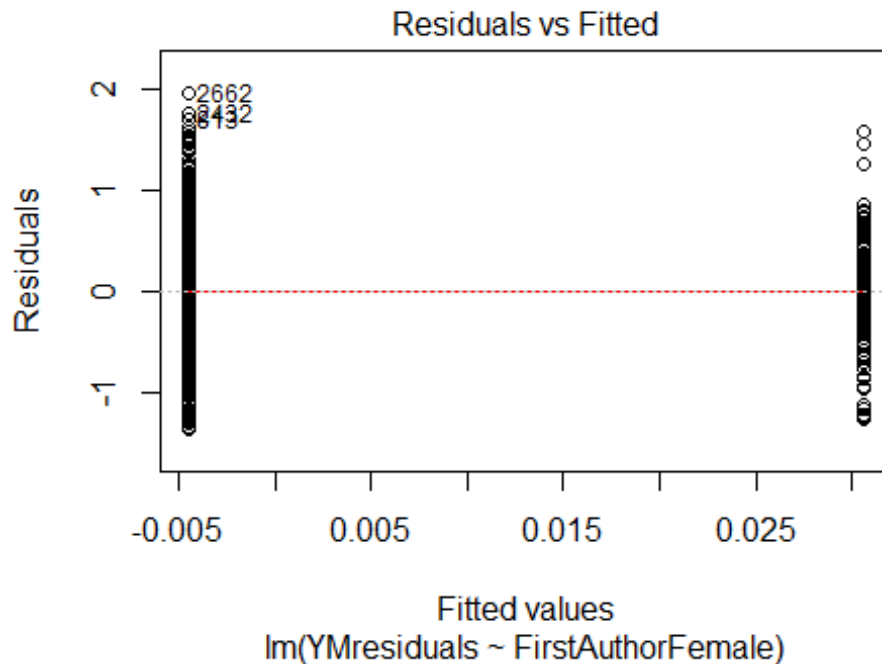
## Year2005          0.0908      0.0648      1.40      0.161
## Year2006          0.0697      0.0605      1.15      0.249
## Year2007          0.1024      0.0588      1.74      0.082 .
## Year2008          0.0392      0.0570      0.69      0.492
## Year2009          0.0699      0.0575      1.22      0.224
## Year2010          0.0827      0.0582      1.42      0.155
## Year2011          0.0550      0.0602      0.91      0.361
## Year2012          0.0482      0.0591      0.82      0.415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.607
## Multiple R-squared:  0.0107, Adjusted R-squared:  0.00675
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 750 is an outlier with |weight| = 0 ( < 2.3e-05);
## 367 weights are ~= 1. The remaining 3912 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0008 0.8700 0.9510 0.9070 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          2.34e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4280"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2605"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   83   70  104   85  135  122  122  115  122  154  152  159  151  185  175
## 2011 2012
##   157  190
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    23    30    49    31    32    39    56    58    58    87    77    86    90   105   96
## 2011 2012
##    93   116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    20    27    49    29    32    37    52    53    52    79    65    72    79    94    90
## 2011 2012
##    84   104
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```

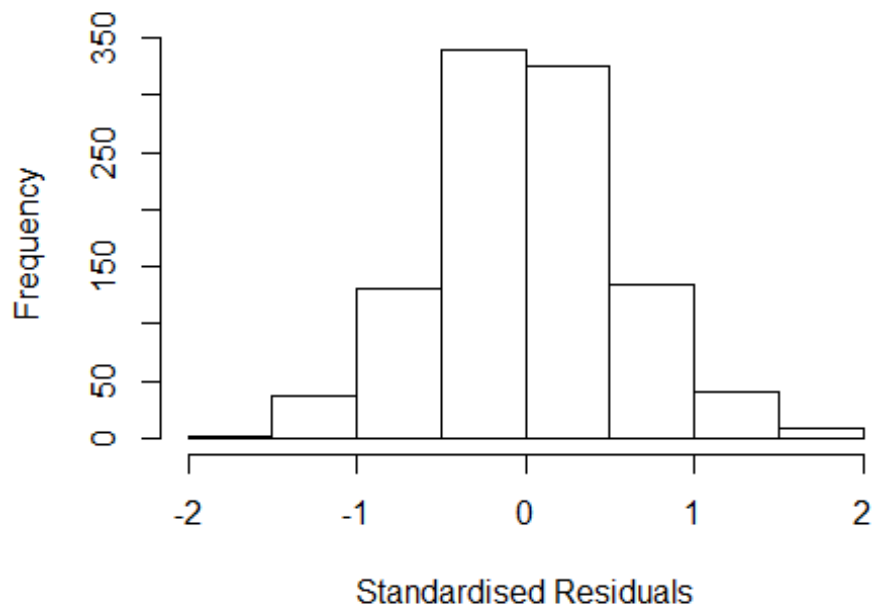


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.2, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 2.65634131315477"
## [1] "Male first author team size 2018 geometric mean: 2.35173305935062"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 790, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.82958041096812"
## [1] "Male last author team size 2018 geometric mean: 2.28881101349504"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.227 1          1.108
## LastAuthorFemale  1.158 1          1.076
## UniqueAuthors    1.553 4          1.057
## Year              1.594 16         1.015
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.530106 -0.352979 0.000494 0.364415 1.986897
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27129 0.13598 9.35 < 2e-16 ***
## FirstAuthorFemale1 -0.01355 0.04933 -0.27 0.7837
## LastAuthorFemale1 0.00500 0.06435 0.08 0.9381
## UniqueAuthors2 0.26206 0.04608 5.69 1.7e-08 ***
## UniqueAuthors3 0.30689 0.05167 5.94 4.0e-09 ***
## UniqueAuthors4 0.26900 0.08163 3.30 0.0010 **
## UniqueAuthors5 0.48588 0.07888 6.16 1.1e-09 ***
## Year1997 -0.10046 0.16711 -0.60 0.5479
## Year1998 -0.20089 0.16881 -1.19 0.2343
## Year1999 -0.00324 0.15271 -0.02 0.9831
```

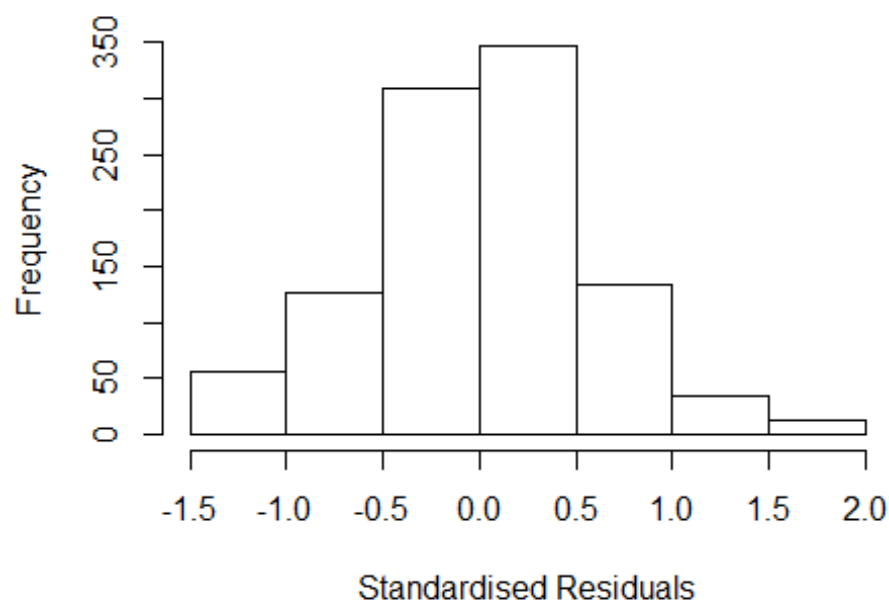
```

## Year2000      -0.11248    0.19065   -0.59    0.5553
## Year2001      -0.13858    0.16725   -0.83    0.4076
## Year2002      -0.29119    0.16761   -1.74    0.0826 .
## Year2003      -0.10799    0.15148   -0.71    0.4761
## Year2004      -0.39952    0.16343   -2.44    0.0147 *
## Year2005      -0.18841    0.14941   -1.26    0.2076
## Year2006      -0.28289    0.14775   -1.91    0.0558 .
## Year2007      -0.43011    0.15472   -2.78    0.0055 **
## Year2008      -0.27748    0.14770   -1.88    0.0606 .
## Year2009      -0.23280    0.14568   -1.60    0.1103
## Year2010      -0.26364    0.14890   -1.77    0.0769 .
## Year2011      -0.28125    0.14868   -1.89    0.0588 .
## Year2012      -0.31682    0.14599   -2.17    0.0302 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.0935, Adjusted R-squared:  0.0735
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 914 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.123  0.859  0.948  0.896  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.137 1      1.066
## LastAuthorFemale  1.103 1      1.050
## Year              1.178 16      1.005

```



## Residuals from first and last author



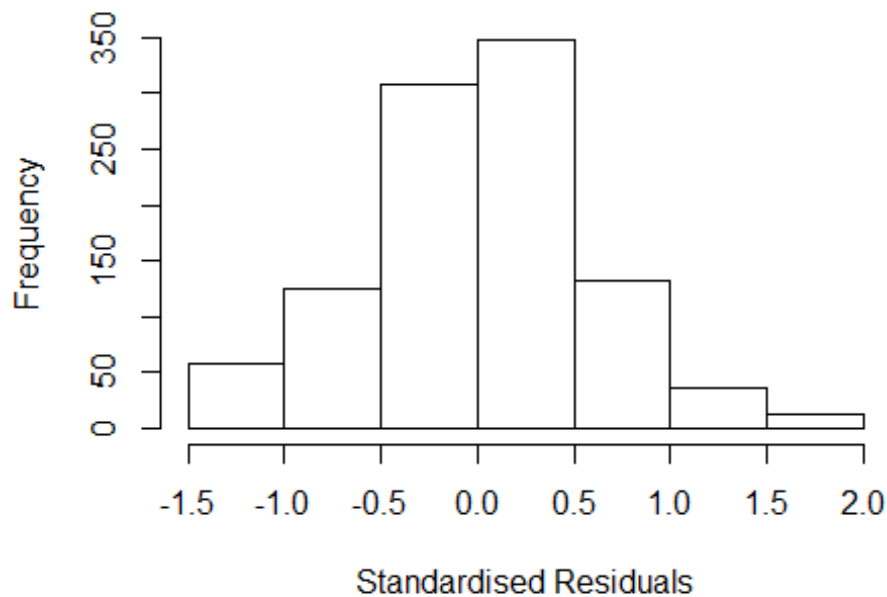
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4315 -0.3544 0.0231 0.3531 1.9651
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36172 0.15465 8.81 <2e-16 ***
## FirstAuthorFemale1 0.06043 0.04895 1.23 0.217
## LastAuthorFemale1 0.00936 0.06478 0.14 0.885
## Year1997 -0.08140 0.18758 -0.43 0.664
## Year1998 -0.16428 0.19238 -0.85 0.393
## Year1999 0.04891 0.17582 0.28 0.781
## Year2000 -0.05125 0.21086 -0.24 0.808
## Year2001 -0.05159 0.18752 -0.28 0.783
## Year2002 -0.15757 0.18741 -0.84 0.401
## Year2003 0.00508 0.16920 0.03 0.976
## Year2004 -0.30263 0.18407 -1.64 0.100
## Year2005 -0.07195 0.16670 -0.43 0.666
```

```

## Year2006      -0.22446    0.16508   -1.36    0.174
## Year2007      -0.30100    0.16986   -1.77    0.077 .
## Year2008      -0.18319    0.16672   -1.10    0.272
## Year2009      -0.10396    0.16495   -0.63    0.529
## Year2010      -0.11216    0.16643   -0.67    0.501
## Year2011      -0.17082    0.16714   -1.02    0.307
## Year2012      -0.19231    0.16500   -1.17    0.244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.0256, Adjusted R-squared:  0.00806
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 942 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.163  0.866  0.952  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.114 1          1.055
## Year              1.114 16          1.003

```

## Residuals from first author

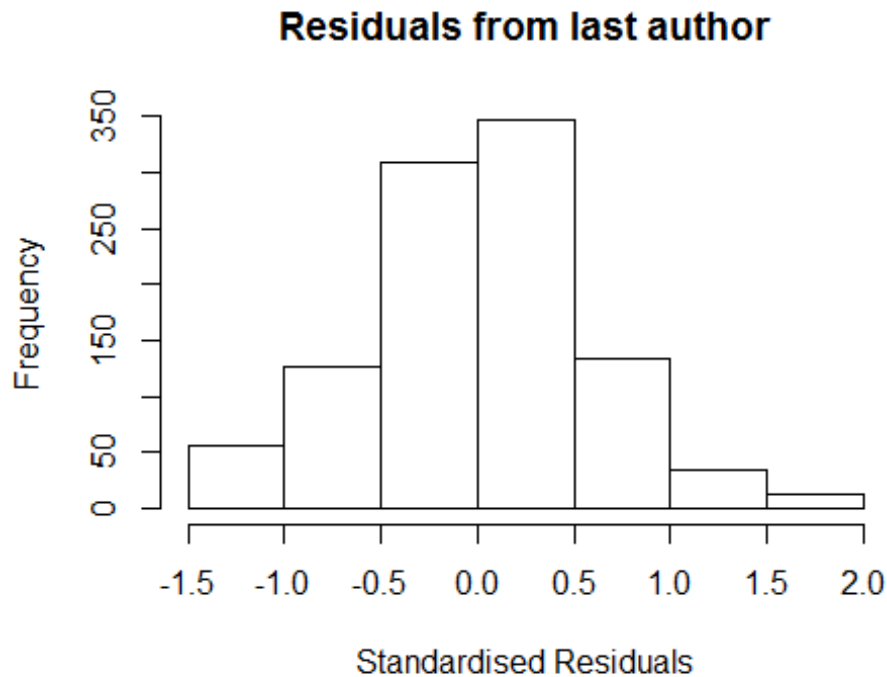


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4241 -0.3499 0.0225 0.3549 1.9649
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36214 0.15457 8.81 <2e-16 ***
## FirstAuthorFemale1 0.06196 0.04841 1.28 0.201
## Year1997 -0.08190 0.18752 -0.44 0.662
## Year1998 -0.16362 0.19169 -0.85 0.394
## Year1999 0.04878 0.17572 0.28 0.781
## Year2000 -0.05012 0.21062 -0.24 0.812
## Year2001 -0.05120 0.18720 -0.27 0.785
## Year2002 -0.15773 0.18705 -0.84 0.399
## Year2003 0.00532 0.16896 0.03 0.975
## Year2004 -0.30276 0.18403 -1.65 0.100
## Year2005 -0.07187 0.16638 -0.43 0.666
## Year2006 -0.22416 0.16476 -1.36 0.174
```

```

## Year2007      -0.30066    0.16943   -1.77    0.076 .
## Year2008      -0.18329    0.16653   -1.10    0.271
## Year2009      -0.10355    0.16451   -0.63    0.529
## Year2010      -0.11205    0.16619   -0.67    0.500
## Year2011      -0.17102    0.16695   -1.02    0.306
## Year2012      -0.19174    0.16440   -1.17    0.244
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.00912
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 942 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.159  0.864  0.952  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.077 1          1.038
## Year            1.077 16          1.002

```



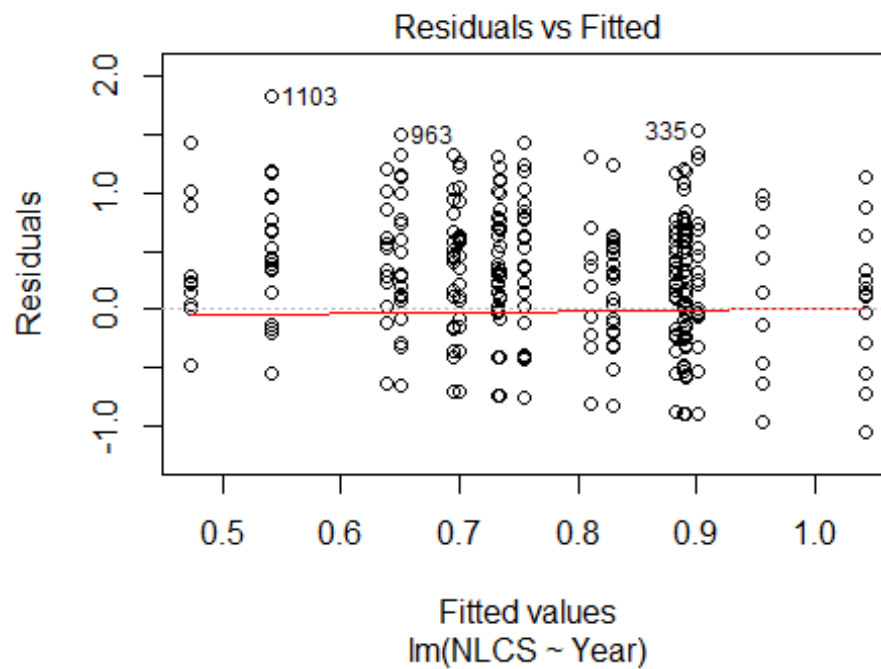
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.412 -0.357 0.022 0.354 1.960
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.36721 0.15466 8.84 <2e-16 ***
## LastAuthorFemale1 0.02300 0.06356 0.36 0.718
## Year1997 -0.07908 0.18785 -0.42 0.674
## Year1998 -0.15484 0.19177 -0.81 0.420
## Year1999 0.04492 0.17606 0.26 0.799
## Year2000 -0.04876 0.21085 -0.23 0.817
## Year2001 -0.05071 0.18742 -0.27 0.787
## Year2002 -0.15810 0.18758 -0.84 0.400
## Year2003 0.00418 0.16915 0.02 0.980
## Year2004 -0.30732 0.18345 -1.68 0.094 .
## Year2005 -0.06679 0.16634 -0.40 0.688
## Year2006 -0.22602 0.16518 -1.37 0.172
```

```

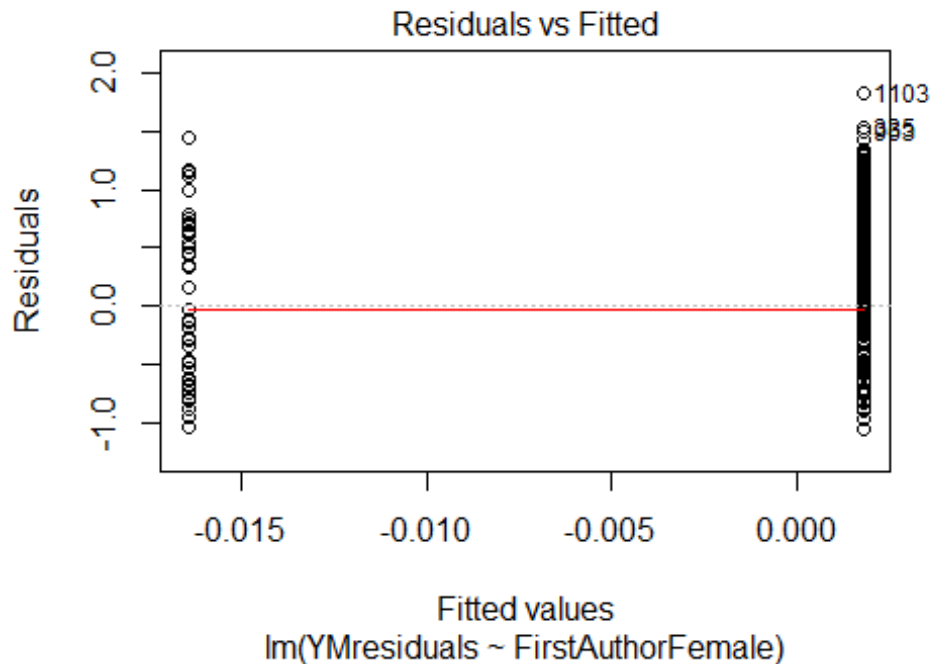
## Year2007          -0.30065      0.16975    -1.77      0.077 .
## Year2008          -0.17761      0.16627    -1.07      0.286
## Year2009          -0.10054      0.16496    -0.61      0.542
## Year2010          -0.11203      0.16632    -0.67      0.501
## Year2011          -0.17086      0.16693    -1.02      0.306
## Year2012          -0.18954      0.16481    -1.15      0.250
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0244, Adjusted R-squared:  0.00783
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 939 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.164  0.863  0.950  0.898  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1018"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2606"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   56   58   64   26   65   71   56   49   37   31   50   65   61   62   66
## 2011 2012
##   66   59
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   22   28   10   14   24   31   27   18   11   31   29   29   43   43
## 2011 2012

```

```
## 42 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 22 28 10 11 23 28 27 17 11 27 29 27 34 40
## 2011 2012
## 38 37
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 6.9, df = 16, p-value = 1
```



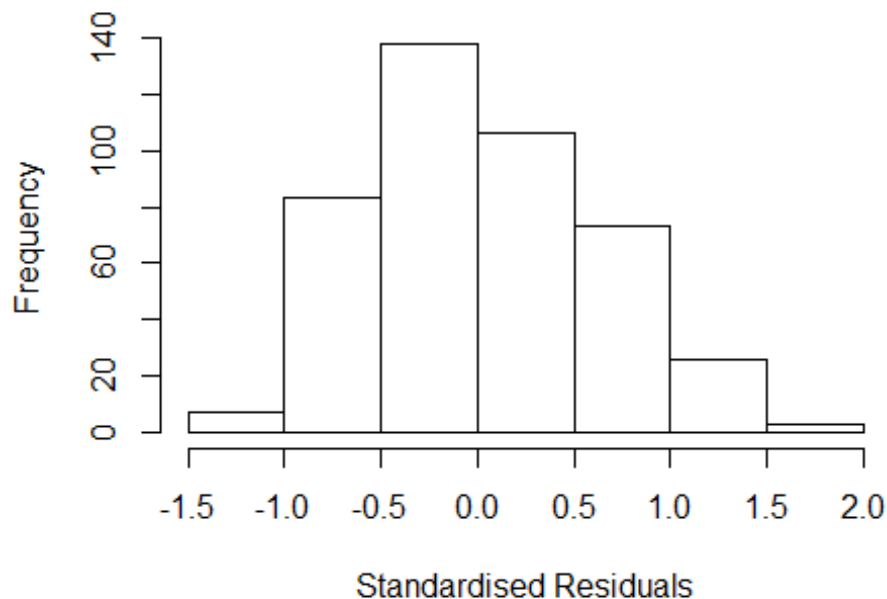
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.65, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.05756580202225"
## [1] "Male first author team size 2018 geometric mean: 2.28430927725531"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 470, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.95598117719593"
## [1] "Male last author team size 2018 geometric mean: 2.46813115907057"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.647  1      1.283
## LastAuthorFemale  1.653  1      1.286
## UniqueAuthors    1.881  4      1.082
## Year             2.350 16      1.027
```



## Residuals from first and last author and team size



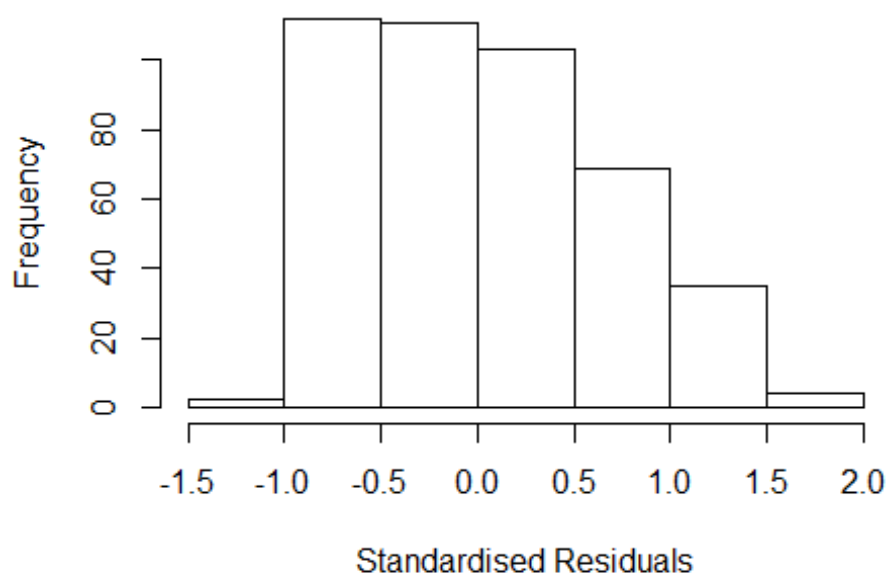
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2871 -0.4301 -0.0343 0.4384 1.7239
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4864 0.1358 3.58 0.00038 ***
## FirstAuthorFemale1 0.0931 0.1335 0.70 0.48601
## LastAuthorFemale1 -0.3019 0.1139 -2.65 0.00832 **
## UniqueAuthors2 0.4131 0.0772 5.35 1.5e-07 ***
## UniqueAuthors3 0.4011 0.1253 3.20 0.00147 **
## UniqueAuthors4 0.2731 0.1712 1.60 0.11139
## UniqueAuthors5 0.1912 0.1869 1.02 0.30692
## Year1997 -0.0264 0.1874 -0.14 0.88807
## Year1998 0.0980 0.1857 0.53 0.59818
## Year1999 0.3172 0.2664 1.19 0.23448
```

```

## Year2000          0.3065      0.2668      1.15  0.25130
## Year2001          0.2575      0.2055      1.25  0.21082
## Year2002         -0.0762      0.1894     -0.40  0.68778
## Year2003          0.2045      0.1769      1.16  0.24829
## Year2004         -0.1391      0.1932     -0.72  0.47203
## Year2005          0.1817      0.2763      0.66  0.51122
## Year2006          0.1283      0.1717      0.75  0.45541
## Year2007          0.0492      0.1771      0.28  0.78121
## Year2008          0.2505      0.1773      1.41  0.15841
## Year2009         -0.1240      0.1669     -0.74  0.45786
## Year2010         -0.0564      0.1497     -0.38  0.70682
## Year2011         -0.1653      0.1611     -1.03  0.30535
## Year2012          0.2136      0.1723      1.24  0.21594
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.148, Adjusted R-squared:  0.103
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 404 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.425  0.874  0.945  0.913  0.976  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.29e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.462 1      1.209
## LastAuthorFemale  1.427 1      1.195
## Year              1.221 16      1.006

```

## Residuals from first and last author

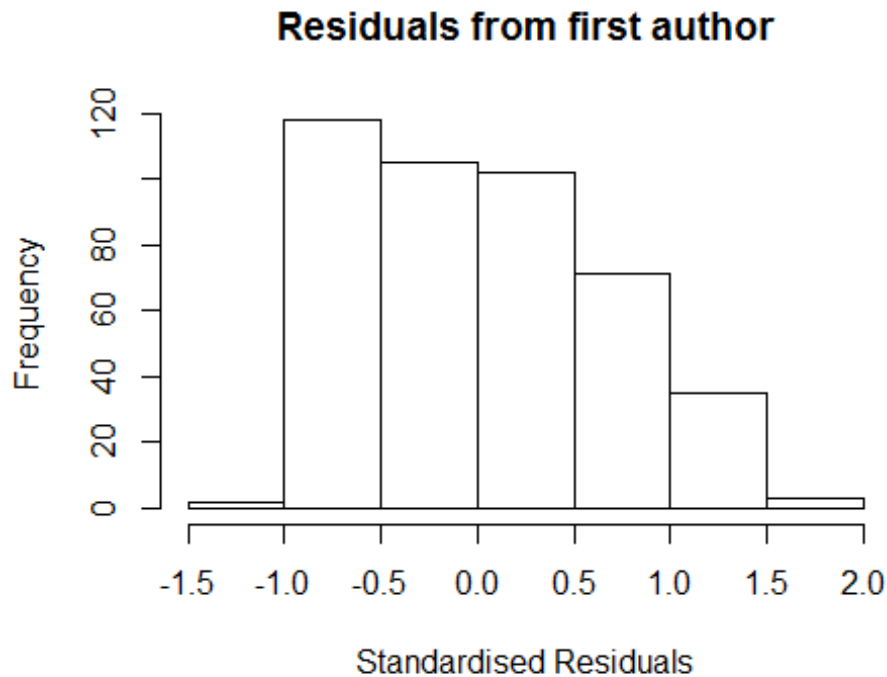


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1855 -0.5452 -0.0235 0.4903 1.9051
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.69064 0.13828 4.99 8.7e-07 ***
## FirstAuthorFemale1 0.07420 0.13248 0.56 0.576
## LastAuthorFemale1 -0.23970 0.10864 -2.21 0.028 *
## Year1997 -0.06199 0.19778 -0.31 0.754
## Year1998 0.00403 0.18621 0.02 0.983
## Year1999 0.27755 0.28818 0.96 0.336
## Year2000 0.42062 0.26417 1.59 0.112
## Year2001 0.18897 0.22921 0.82 0.410
## Year2002 -0.07815 0.20834 -0.38 0.708
## Year2003 0.21018 0.18101 1.16 0.246
## Year2004 -0.20766 0.19689 -1.05 0.292
## Year2005 0.14146 0.25071 0.56 0.573
```

```

## Year2006          0.07984    0.16997    0.47    0.639
## Year2007          0.03200    0.18381    0.17    0.862
## Year2008          0.20220    0.19548    1.03    0.302
## Year2009         -0.14356    0.18339   -0.78    0.434
## Year2010         -0.08716    0.16358   -0.53    0.594
## Year2011         -0.23073    0.16846   -1.37    0.172
## Year2012          0.18845    0.17945    1.05    0.294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.664
## Multiple R-squared:  0.0657, Adjusted R-squared:  0.0253
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 411 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.390  0.895   0.939   0.917   0.979   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.29e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.104 1      1.051
## Year              1.104 16      1.003

```



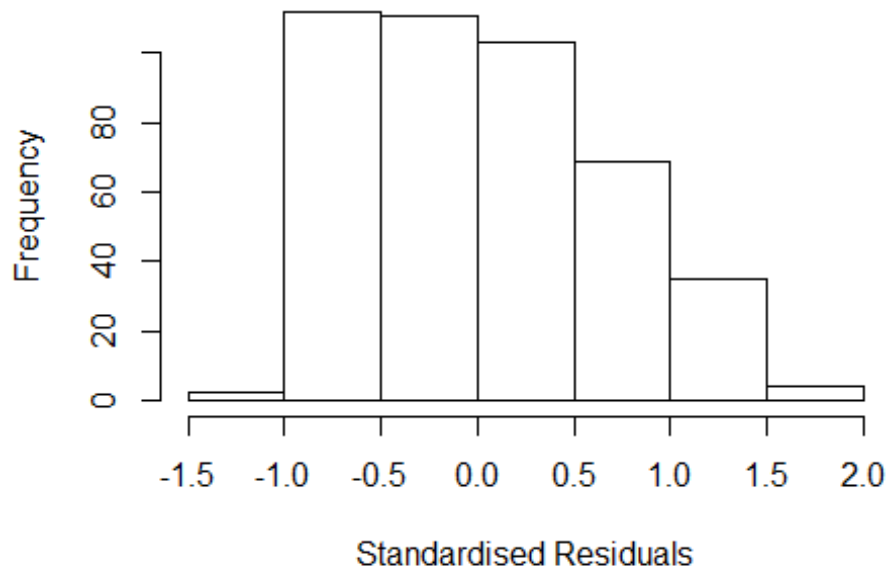
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0796 -0.5452 -0.0393  0.4984  1.9171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.67749    0.13626   4.97 9.7e-07 ***
## FirstAuthorFemale1 -0.04456    0.12148  -0.37  0.71
## Year1997        -0.05542    0.19728  -0.28  0.78
## Year1998        -0.00235    0.18799  -0.01  0.99
## Year1999         0.28117    0.29330   0.96  0.34
## Year2000         0.40210    0.25800   1.56  0.12
## Year2001         0.19247    0.22797   0.84  0.40
## Year2002        -0.08139    0.20490  -0.40  0.69
## Year2003         0.21758    0.18563   1.17  0.24
## Year2004        -0.22010    0.19560  -1.13  0.26
## Year2005         0.11427    0.25029   0.46  0.65
## Year2006         0.07333    0.16887   0.43  0.66
```

```

## Year2007          0.05146    0.18204    0.28    0.78
## Year2008          0.18961    0.19120    0.99    0.32
## Year2009         -0.13229    0.18330   -0.72    0.47
## Year2010         -0.10240    0.16266   -0.63    0.53
## Year2011         -0.22962    0.16627   -1.38    0.17
## Year2012          0.16595    0.17713    0.94    0.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.663
## Multiple R-squared:  0.0561, Adjusted R-squared:  0.0177
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 29 weights are ~= 1. The remaining 407 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.383  0.888  0.939  0.915  0.978  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.29e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.078 1          1.038
## Year            1.078 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1093 -0.5297 -0.0292 0.4877 1.8977
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.696391 0.135748 5.13 4.4e-07 ***
## LastAuthorFemale1 -0.208648 0.093453 -2.23 0.026 *
## Year1997 -0.066423 0.195638 -0.34 0.734
## Year1998 0.000118 0.185132 0.00 0.999
## Year1999 0.275504 0.284252 0.97 0.333
## Year2000 0.412901 0.258414 1.60 0.111
## Year2001 0.191237 0.230696 0.83 0.408
## Year2002 -0.083698 0.207665 -0.40 0.687
## Year2003 0.207729 0.181379 1.15 0.253
## Year2004 -0.213019 0.194473 -1.10 0.274
## Year2005 0.133936 0.245479 0.55 0.586
## Year2006 0.076034 0.167844 0.45 0.651
```

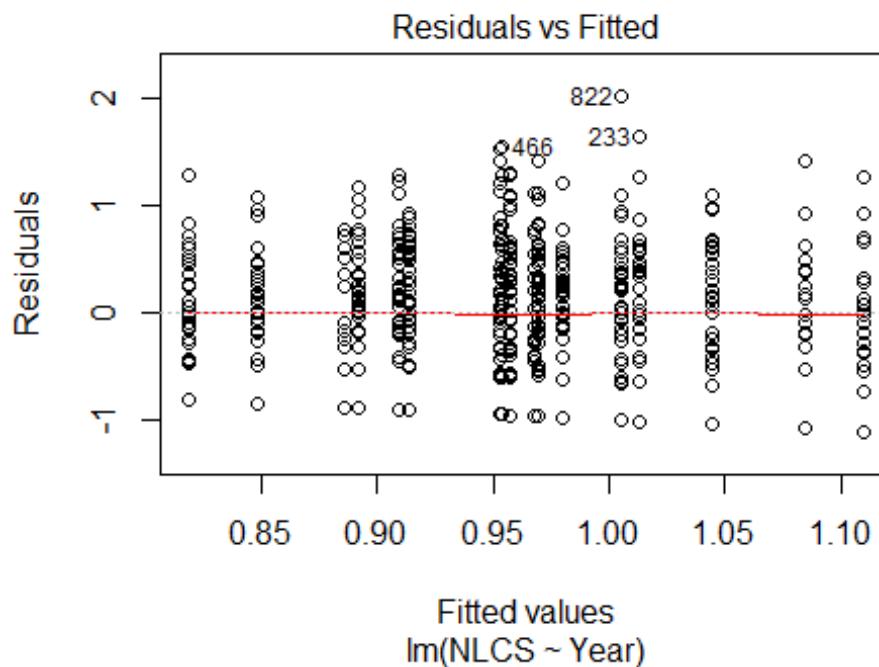
```

## Year2007      0.033244    0.183843    0.18    0.857
## Year2008      0.195592    0.193537    1.01    0.313
## Year2009     -0.142666    0.182372   -0.78    0.434
## Year2010     -0.087228    0.162271   -0.54    0.591
## Year2011     -0.229109    0.167660   -1.37    0.173
## Year2012      0.181136    0.176326    1.03    0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.0651, Adjusted R-squared:  0.0271
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 408 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.389  0.892  0.937  0.916  0.979  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.29e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 436"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2607"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   37   63   32   56   50   54   55   42   48   28   48   50   67   73   68
## 2011 2012
##   76   57
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   34   19   32   20   37   35   28   32   19   36   36   48   53   52
## 2011 2012

```



```
## 57 42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 33 17 32 20 36 34 25 29 19 35 32 44 50 47
## 2011 2012
## 55 42
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 13, df = 16, p-value = 0.7
```



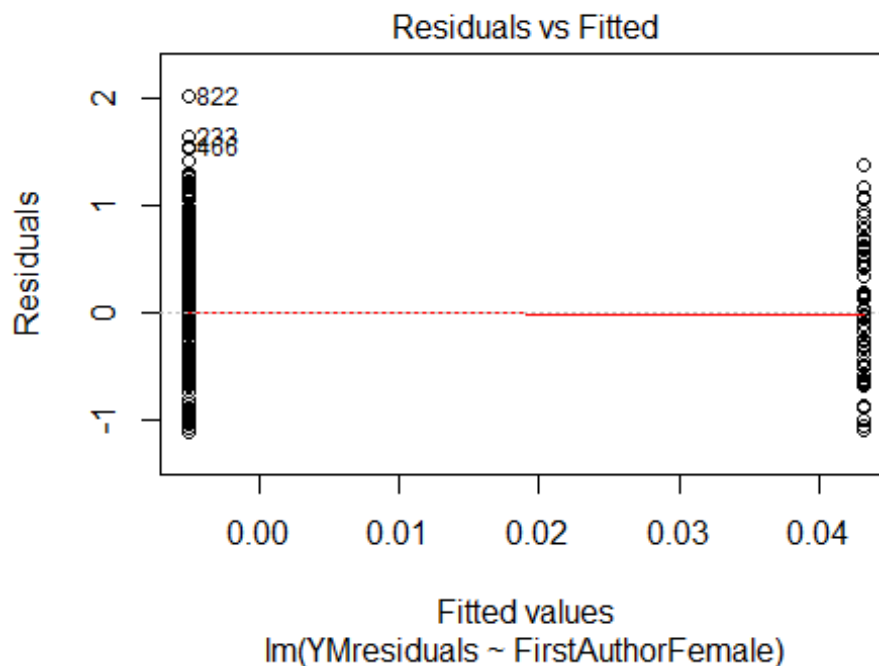
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.013, df = 1, p-value = 0.9

## [1] "Female first author team size 2018 geometric mean: 2.1192680453341"
## [1] "Male first author team size 2018 geometric mean: 1.52992054608469"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

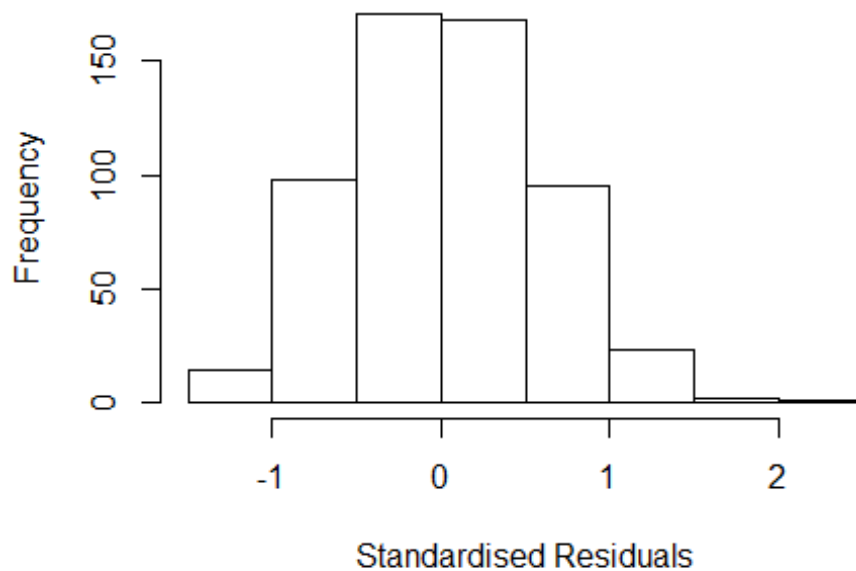
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.44948974278318"
## [1] "Male last author team size 2018 geometric mean: 1.57721802399189"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 69, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.780  1      1.334
## LastAuthorFemale  1.818  1      1.348
## UniqueAuthors    1.385  4      1.042
## Year              1.591 16      1.015
```

## Residuals from first and last author and team size



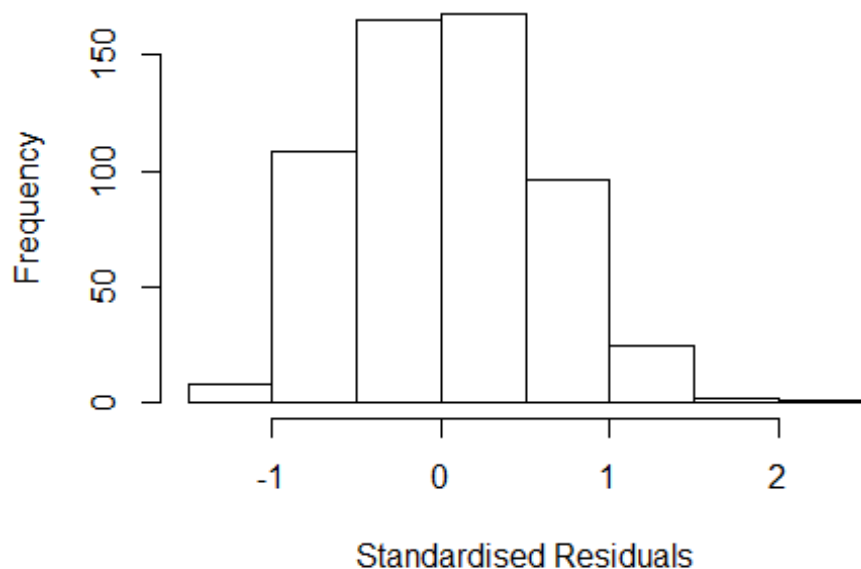
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18770 -0.38676 0.00278 0.42343 2.05382
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0661 0.1112 9.59 < 2e-16 ***
## FirstAuthorFemale1 0.0379 0.1058 0.36 0.72023
## LastAuthorFemale1 0.0212 0.1039 0.20 0.83832
## UniqueAuthors2 0.1808 0.0611 2.96 0.00320 **
## UniqueAuthors3 0.0681 0.1088 0.63 0.53181
## UniqueAuthors4 0.5311 0.1421 3.74 0.00021 ***
## UniqueAuthors5 0.6026 0.1909 3.16 0.00168 **
## Year1997 -0.1256 0.1436 -0.87 0.38232
## Year1998 -0.1872 0.1789 -1.05 0.29585
## Year1999 -0.0926 0.1768 -0.52 0.60055
```

```

## Year2000          -0.0407      0.1730   -0.24   0.81426
## Year2001          -0.2484      0.1552   -1.60   0.11017
## Year2002          -0.2249      0.1583   -1.42   0.15606
## Year2003          -0.2164      0.1868   -1.16   0.24720
## Year2004          -0.3669      0.1517   -2.42   0.01593 *
## Year2005          -0.2038      0.1759   -1.16   0.24714
## Year2006          -0.0971      0.1490   -0.65   0.51483
## Year2007          -0.3729      0.1486   -2.51   0.01238 *
## Year2008          -0.1740      0.1445   -1.20   0.22896
## Year2009          -0.2083      0.1447   -1.44   0.15070
## Year2010          -0.2030      0.1368   -1.48   0.13823
## Year2011          -0.1922      0.1407   -1.37   0.17256
## Year2012          -0.2101      0.1538   -1.37   0.17264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.0487, Adjusted R-squared:  0.0105
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 527 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.871  0.950  0.913  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.75e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.695 1          1.302
## LastAuthorFemale  1.742 1          1.320
## Year              1.194 16          1.006

```

## Residuals from first and last author



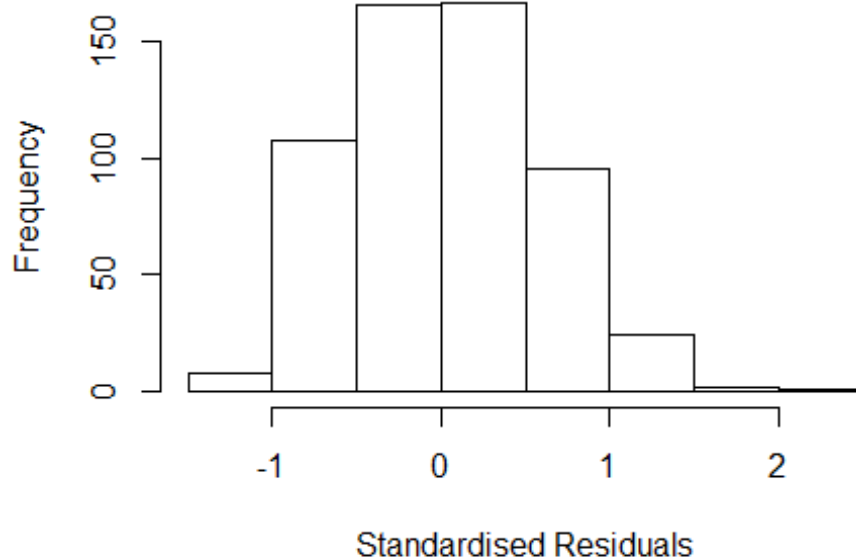
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10273 -0.39972 0.00547 0.41311 2.02521
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0935 0.1134 9.64 <2e-16 ***
## FirstAuthorFemale1 0.0472 0.1071 0.44 0.660
## LastAuthorFemale1 0.0175 0.1050 0.17 0.868
## Year1997 -0.1142 0.1446 -0.79 0.430
## Year1998 -0.1513 0.1774 -0.85 0.394
## Year1999 -0.0970 0.1788 -0.54 0.588
## Year2000 -0.0269 0.1751 -0.15 0.878
## Year2001 -0.2260 0.1600 -1.41 0.158
## Year2002 -0.1983 0.1586 -1.25 0.212
## Year2003 -0.1898 0.1891 -1.00 0.316
## Year2004 -0.3192 0.1530 -2.09 0.037 *
## Year2005 -0.2110 0.1790 -1.18 0.239
```

```

## Year2006          -0.0379      0.1557   -0.24    0.808
## Year2007          -0.2988      0.1464   -2.04    0.042 *
## Year2008          -0.1047      0.1482   -0.71    0.480
## Year2009          -0.1678      0.1477   -1.14    0.256
## Year2010          -0.1684      0.1385   -1.22    0.225
## Year2011          -0.1303      0.1437   -0.91    0.365
## Year2012          -0.1918      0.1536   -1.25    0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0176, Adjusted R-squared:  -0.0144
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 520 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.239  0.856  0.948  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.75e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.086 1      1.042
## Year              1.086 16      1.003

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11166 -0.40335 0.00688 0.41297 2.02383
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0935 0.1134 9.64 <2e-16 ***
## FirstAuthorFemale1 0.0555 0.0871 0.64 0.524
## Year1997 -0.1129 0.1445 -0.78 0.435
## Year1998 -0.1518 0.1775 -0.86 0.393
## Year1999 -0.0954 0.1780 -0.54 0.592
## Year2000 -0.0260 0.1757 -0.15 0.882
## Year2001 -0.2254 0.1598 -1.41 0.159
## Year2002 -0.1975 0.1586 -1.24 0.214
## Year2003 -0.1883 0.1885 -1.00 0.318
## Year2004 -0.3170 0.1518 -2.09 0.037 *
## Year2005 -0.2095 0.1788 -1.17 0.242
## Year2006 -0.0374 0.1555 -0.24 0.810
```

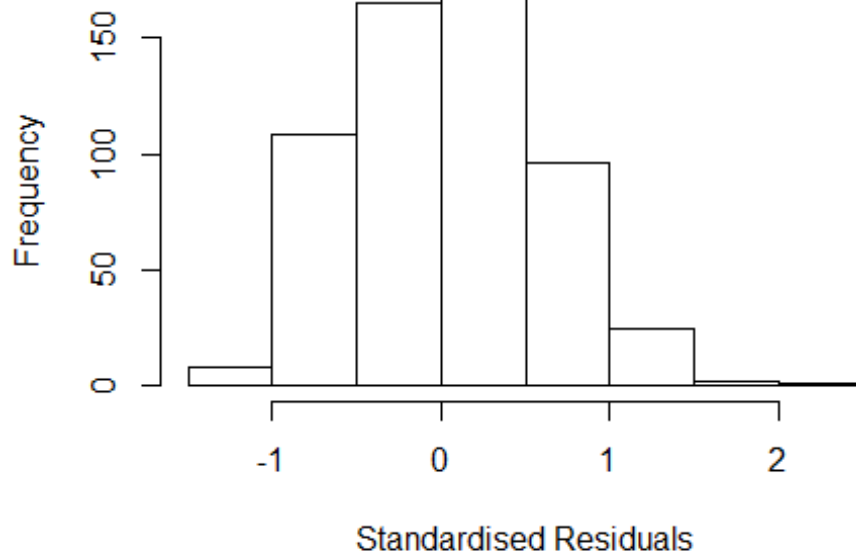
```

## Year2007          -0.2970      0.1472   -2.02    0.044 *
## Year2008          -0.1034      0.1481   -0.70    0.486
## Year2009          -0.1668      0.1479   -1.13    0.260
## Year2010          -0.1678      0.1384   -1.21    0.226
## Year2011          -0.1304      0.1436   -0.91    0.364
## Year2012          -0.1918      0.1535   -1.25    0.212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.605
## Multiple R-squared:  0.0176, Adjusted R-squared:  -0.0126
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 518 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.241  0.855   0.947   0.910   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.75e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.126 1          1.061
## Year            1.126 16          1.004

```



## Residuals from last author



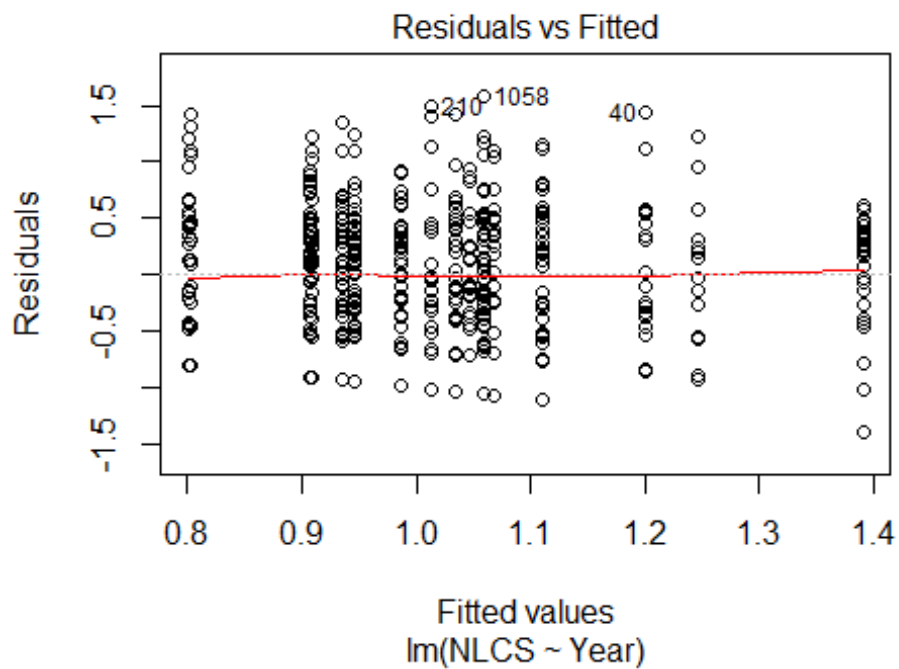
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.09346 -0.39657 0.00962 0.41518 2.01860
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0935 0.1134 9.64 <2e-16 ***
## LastAuthorFemale1 0.0409 0.0845 0.48 0.629
## Year1997 -0.1132 0.1449 -0.78 0.435
## Year1998 -0.1483 0.1766 -0.84 0.401
## Year1999 -0.0960 0.1786 -0.54 0.591
## Year2000 -0.0264 0.1746 -0.15 0.880
## Year2001 -0.2234 0.1606 -1.39 0.165
## Year2002 -0.1994 0.1584 -1.26 0.209
## Year2003 -0.1870 0.1886 -0.99 0.322
## Year2004 -0.3200 0.1524 -2.10 0.036 *
## Year2005 -0.2068 0.1788 -1.16 0.248
## Year2006 -0.0309 0.1559 -0.20 0.843
```

```

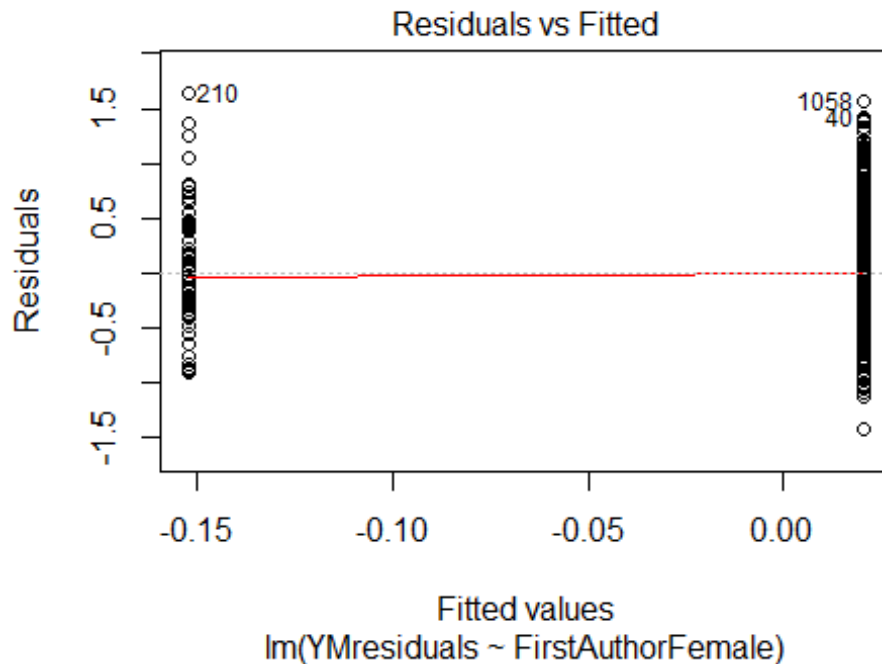
## Year2007          -0.2941      0.1453   -2.02      0.043 *
## Year2008          -0.0981      0.1472   -0.67      0.506
## Year2009          -0.1664      0.1477   -1.13      0.260
## Year2010          -0.1655      0.1387   -1.19      0.233
## Year2011          -0.1266      0.1435   -0.88      0.378
## Year2012          -0.1878      0.1538   -1.22      0.223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.0172, Adjusted R-squared:  -0.013
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 517 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.241  0.856  0.946  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.75e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 572"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2608"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   45   39   39   43   43   46   42   41   47   28   48   67   48   80   68
## 2011 2012
##   95   89
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22   15   24   24   22   24   28   22   29   12   31   44   38   54   50
## 2011 2012

```

```
## 69 64
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 15 20 22 21 24 28 22 28 12 30 42 37 50 48
## 2011 2012
## 61 59
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



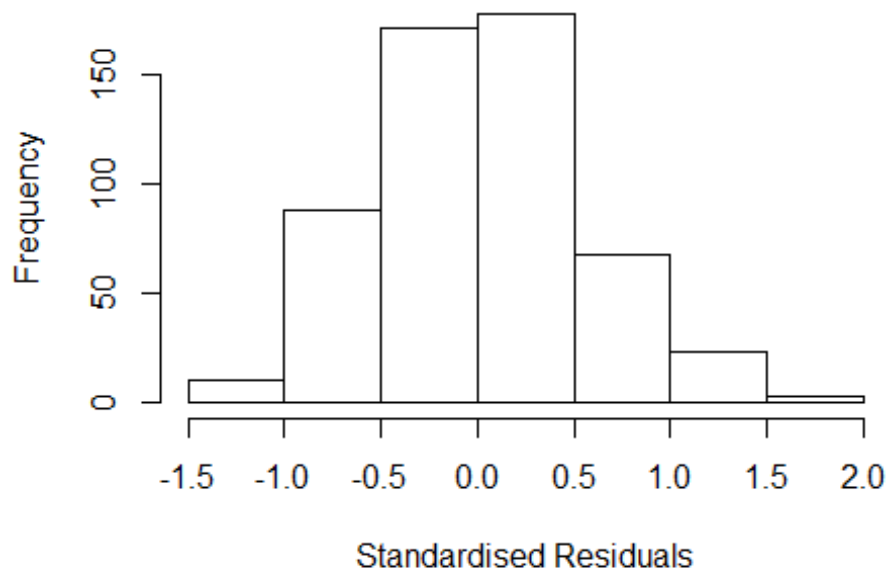
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.2, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 1.68179283050743"
## [1] "Male first author team size 2018 geometric mean: 1.35623981181365"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.5157165665104"
## [1] "Male last author team size 2018 geometric mean: 1.38821936772786"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	4.451	1	2.110
LastAuthorFemale	4.267	1	2.066
UniqueAuthors	14.933	4	1.402
Year	17.797	16	1.094

## Residuals from first and last author and team size



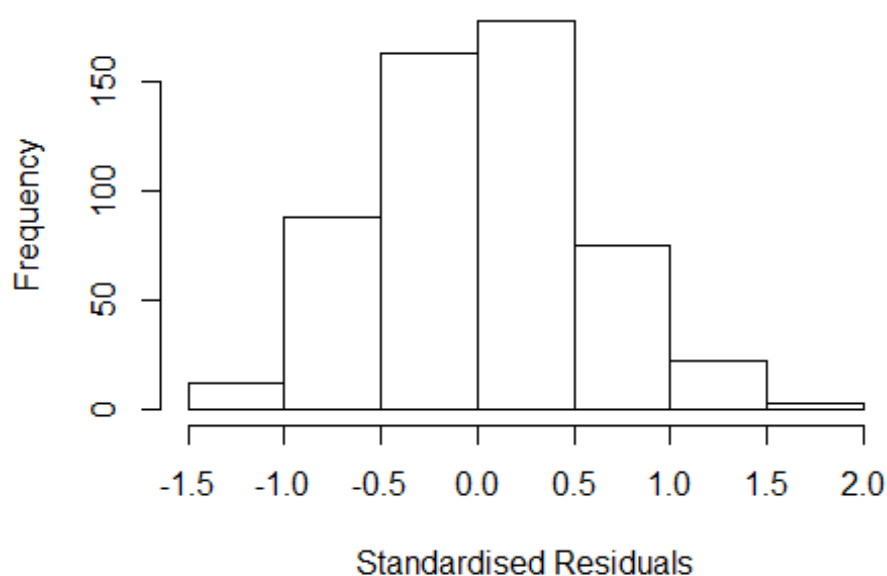
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38347 -0.38535 0.00255 0.38529 1.90359
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1273 0.1361 8.28 1e-15 ***
## FirstAuthorFemale1 -0.0737 0.1359 -0.54 0.58798
## LastAuthorFemale1 -0.2034 0.1410 -1.44 0.14976
## UniqueAuthors2 0.1970 0.0577 3.41 0.00069 ***
## UniqueAuthors3 0.2766 0.1780 1.55 0.12089
## UniqueAuthors4 0.0457 0.1683 0.27 0.78616
## UniqueAuthors5 0.8118 0.0786 10.33 < 2e-16 ***
## Year1997 0.0832 0.2116 0.39 0.69445
## Year1998 -0.1578 0.1821 -0.87 0.38671
## Year1999 0.1235 0.2018 0.61 0.54071
```

```

## Year2000          -0.2558      0.1781    -1.44    0.15160
## Year2001          -0.0636      0.1903    -0.33    0.73840
## Year2002          -0.2176      0.1793    -1.21    0.22555
## Year2003          -0.3492      0.2190    -1.59    0.11148
## Year2004           0.2562      0.1655     1.55    0.12219
## Year2005          -0.0337      0.1950    -0.17    0.86296
## Year2006          -0.3974      0.1840    -2.16    0.03121 *
## Year2007          -0.2190      0.1533    -1.43    0.15372
## Year2008          -0.1220      0.1662    -0.73    0.46333
## Year2009          -0.2626      0.1565    -1.68    0.09398 .
## Year2010          -0.2602      0.1617    -1.61    0.10823
## Year2011          -0.1310      0.1533    -0.85    0.39334
## Year2012          -0.2369      0.1490    -1.59    0.11233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.543
## Multiple R-squared:  0.112, Adjusted R-squared:  0.0743
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 497 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.193  0.870  0.949  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.85e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.869 1          1.967
## LastAuthorFemale  3.711 1          1.926
## Year              1.318 16          1.009

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4349 -0.3869 0.0279 0.3971 1.8366
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1771 0.1316 8.94 <2e-16 ***
## FirstAuthorFemale1 -0.0107 0.1392 -0.08 0.939
## LastAuthorFemale1 -0.2491 0.1467 -1.70 0.090 .
## Year1997 0.0711 0.2134 0.33 0.739
## Year1998 -0.1412 0.1764 -0.80 0.424
## Year1999 0.0690 0.1983 0.35 0.728
## Year2000 -0.2559 0.1822 -1.40 0.161
## Year2001 -0.0360 0.1971 -0.18 0.855
## Year2002 -0.2386 0.1777 -1.34 0.180
## Year2003 -0.3896 0.2132 -1.83 0.068 .
## Year2004 0.2578 0.1617 1.59 0.111
## Year2005 -0.0428 0.2054 -0.21 0.835
```

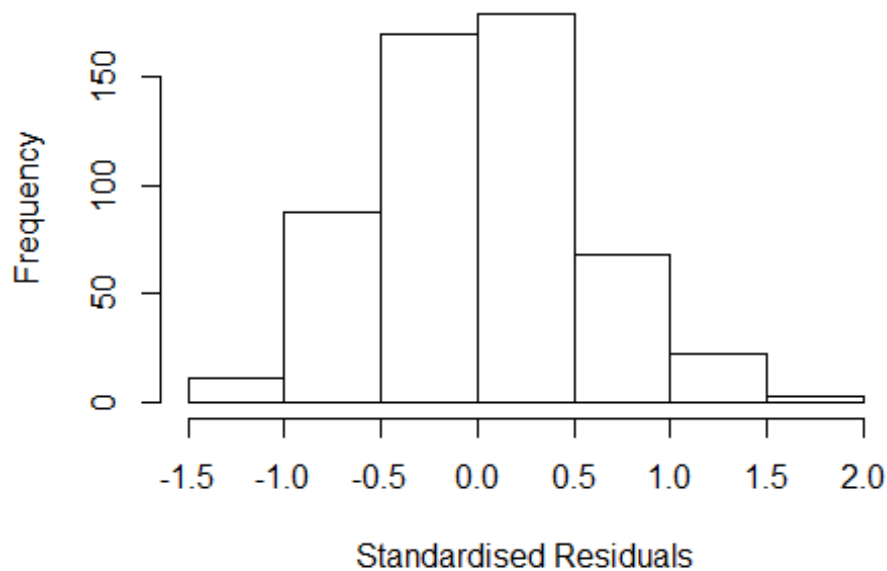
```

## Year2006          -0.3725      0.1874    -1.99      0.047 *
## Year2007          -0.1842      0.1506    -1.22      0.222
## Year2008          -0.1389      0.1627    -0.85      0.393
## Year2009          -0.2688      0.1532    -1.75      0.080 .
## Year2010          -0.2411      0.1581    -1.52      0.128
## Year2011          -0.1270      0.1517    -0.84      0.403
## Year2012          -0.2273      0.1450    -1.57      0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0862, Adjusted R-squared:  0.0547
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 503 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.244  0.872  0.947  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.175 1      1.084
## Year              1.175 16      1.005

```



## Residuals from first author



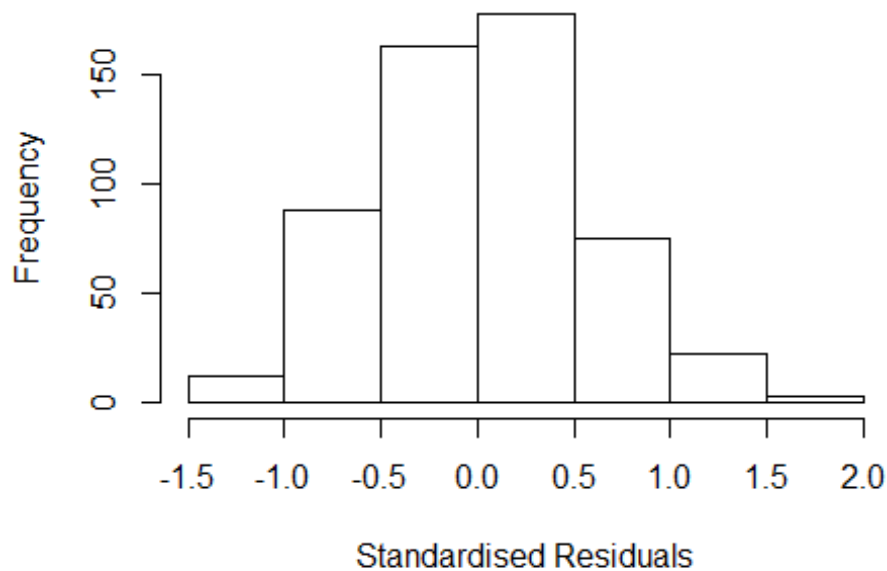
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44013 -0.38990  0.00436  0.38336  1.76765
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1515     0.1283   8.97 <2e-16 ***
## FirstAuthorFemale1 -0.1807     0.0838  -2.15   0.032 *
## Year1997          0.0997     0.2152   0.46   0.644
## Year1998         -0.1191     0.1729  -0.69   0.491
## Year1999          0.0804     0.1975   0.41   0.684
## Year2000         -0.2404     0.1852  -1.30   0.195
## Year2001         -0.0328     0.1953  -0.17   0.867
## Year2002         -0.2298     0.1774  -1.30   0.196
## Year2003         -0.3755     0.2089  -1.80   0.073 .
## Year2004          0.2887     0.1576   1.83   0.068 .
## Year2005         -0.0238     0.1984  -0.12   0.905
## Year2006         -0.3338     0.1890  -1.77   0.078 .
```

```

## Year2007          -0.1627      0.1496    -1.09     0.278
## Year2008          -0.1060      0.1585    -0.67     0.504
## Year2009          -0.2478      0.1506    -1.65     0.101
## Year2010          -0.2281      0.1557    -1.46     0.144
## Year2011          -0.1028      0.1487    -0.69     0.490
## Year2012          -0.2061      0.1422    -1.45     0.148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0773, Adjusted R-squared:  0.0473
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 502 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.292  0.872   0.947   0.907   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.12 1      1.058
## Year              1.12 16      1.004

```

## Residuals from last author



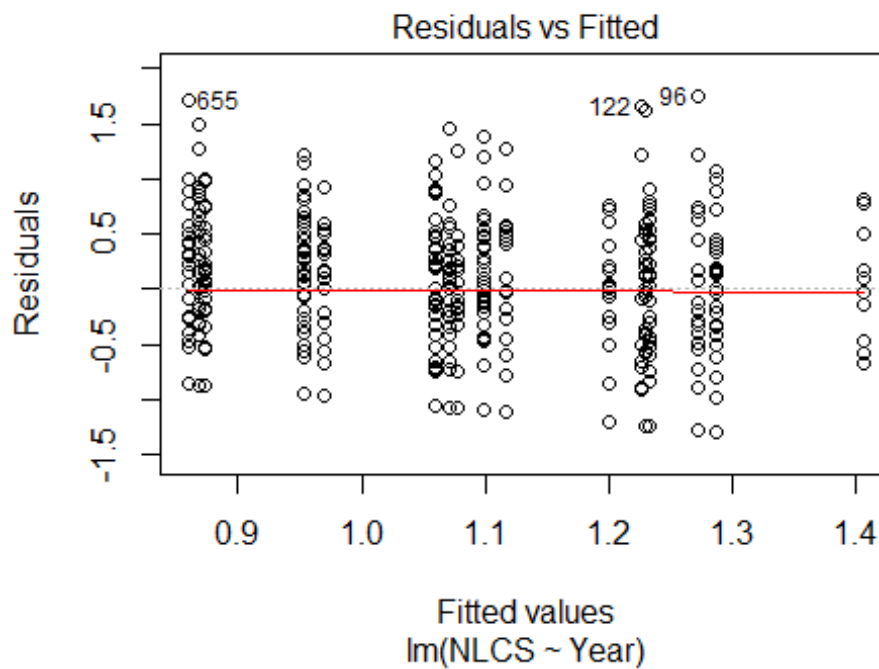
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4350 -0.3864 0.0288 0.3970 1.8358
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1776 0.1311 8.98 <2e-16 ***
## LastAuthorFemale1 -0.2577 0.0825 -3.12 0.0019 **
## Year1997 0.0696 0.2119 0.33 0.7428
## Year1998 -0.1420 0.1759 -0.81 0.4199
## Year1999 0.0687 0.1978 0.35 0.7286
## Year2000 -0.2576 0.1812 -1.42 0.1556
## Year2001 -0.0360 0.1976 -0.18 0.8557
## Year2002 -0.2389 0.1772 -1.35 0.1781
## Year2003 -0.3922 0.2102 -1.87 0.0626 .
## Year2004 0.2575 0.1625 1.58 0.1137
## Year2005 -0.0451 0.1994 -0.23 0.8210
## Year2006 -0.3743 0.1856 -2.02 0.0443 *
```

```

## Year2007          -0.1852      0.1502   -1.23    0.2182
## Year2008          -0.1405      0.1609   -0.87    0.3832
## Year2009          -0.2695      0.1523   -1.77    0.0775 .
## Year2010          -0.2415      0.1578   -1.53    0.1265
## Year2011          -0.1284      0.1497   -0.86    0.3914
## Year2012          -0.2282      0.1440   -1.58    0.1136
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0866, Adjusted R-squared:  0.0569
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 503 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.240  0.871  0.947  0.905  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 541"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2609"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   31   29   19   28   24   25   31   32   35   46   47   57   54   43
## 2011 2012
##   44   42
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   21   20   10   18   11   19   22   25   29   33   39   45   41   32
## 2011 2012

```

```
## 28 36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 21 19 9 18 11 18 22 25 27 31 35 43 41 31
## 2011 2012
## 26 35
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 10, df = 16, p-value = 0.9
```



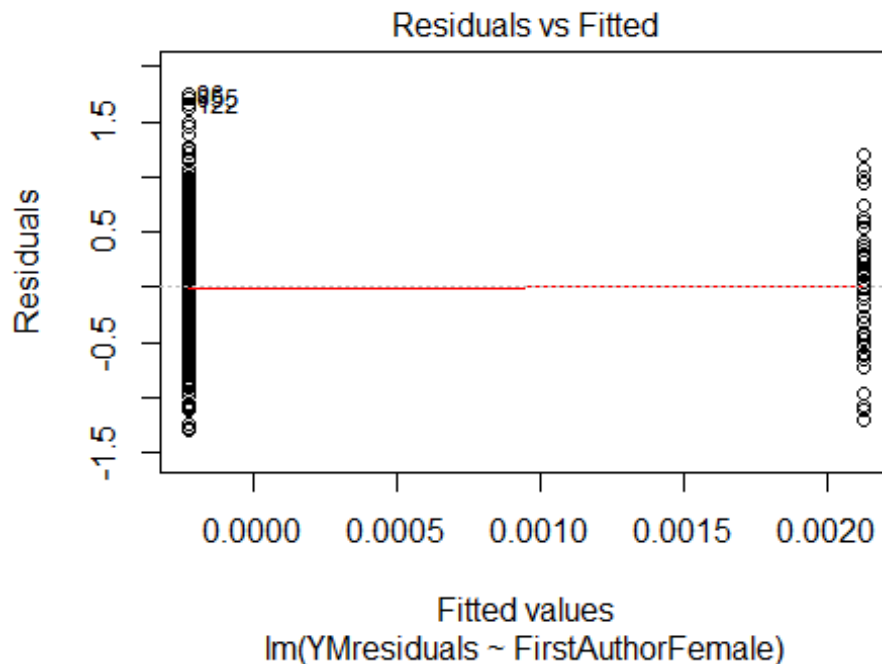
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0025, df = 1, p-value = 1

## [1] "Female first author team size 2018 geometric mean: 1.18920711500272"
## [1] "Male first author team size 2018 geometric mean: 1.07569058622018"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.18920711500272"
## [1] "Male last author team size 2018 geometric mean: 1.07569058622018"

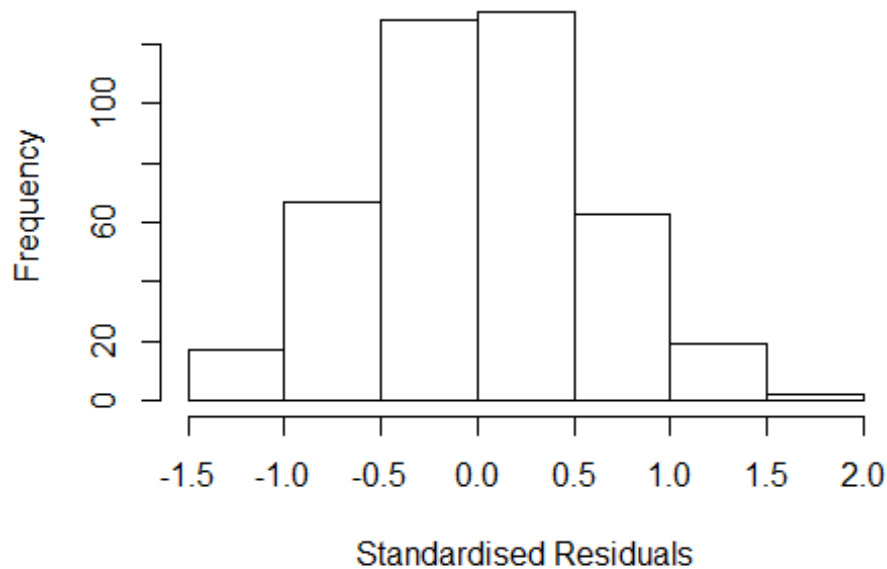
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.664	1	1.290
LastAuthorFemale	1.411	1	1.188
UniqueAuthors	1.677	3	1.090
Year	1.738	16	1.017

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3747 -0.3891 0.0116 0.4011 1.7216
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00551 0.12003 8.38 9e-16 ***
## FirstAuthorFemale1 -0.08503 0.10952 -0.78 0.43795
## LastAuthorFemale1 -0.17062 0.13399 -1.27 0.20362
## UniqueAuthors2 0.20214 0.07578 2.67 0.00795 **
## UniqueAuthors3 0.63819 0.16345 3.90 0.00011 ***
## UniqueAuthors4 0.37278 0.21233 1.76 0.07990 .
## Year1997 0.11383 0.19404 0.59 0.55777
## Year1998 0.16704 0.17768 0.94 0.34770
## Year1999 0.17725 0.36264 0.49 0.62526
## Year2000 0.22226 0.16110 1.38 0.16847
```

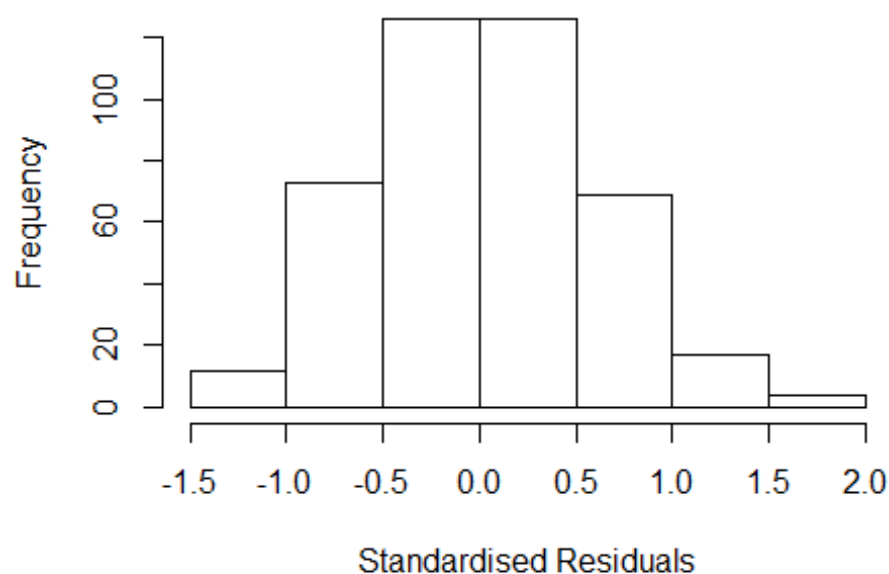
```

## Year2001      0.28609      0.18650      1.53      0.12580
## Year2002     -0.13092      0.16607     -0.79      0.43097
## Year2003      0.12690      0.15654      0.81      0.41803
## Year2004      0.00423      0.15286      0.03      0.97792
## Year2005      0.07155      0.16230      0.44      0.65954
## Year2006     -0.15298      0.16479     -0.93      0.35381
## Year2007     -0.08329      0.15765     -0.53      0.59757
## Year2008     -0.10912      0.15625     -0.70      0.48534
## Year2009      0.04715      0.16030      0.29      0.76880
## Year2010     -0.19609      0.16720     -1.17      0.24159
## Year2011      0.14910      0.17042      0.87      0.38214
## Year2012     -0.20347      0.16997     -1.20      0.23198
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0678
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 391 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.424  0.880  0.954  0.918  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.480 1      1.217
## LastAuthorFemale  1.398 1      1.182
## Year              1.221 16      1.006

```



## Residuals from first and last author



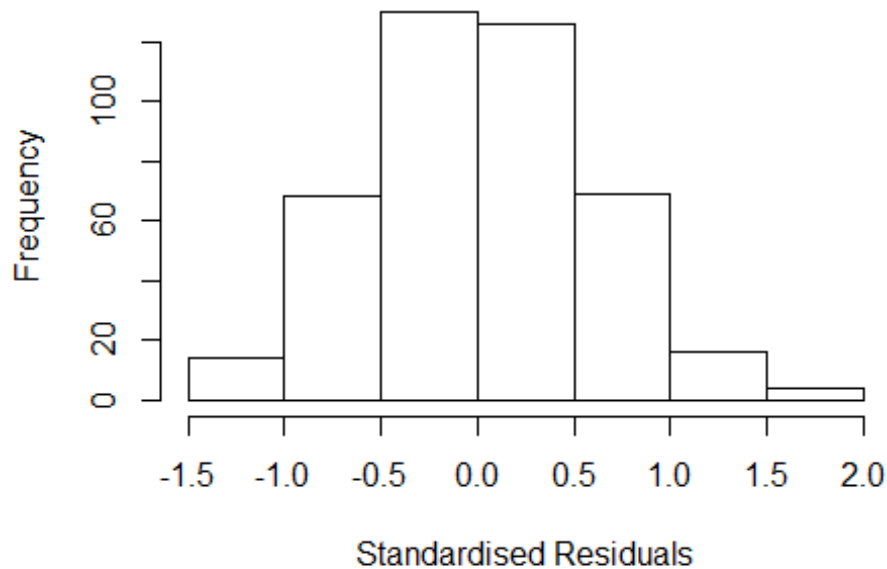
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.249 -0.424 0.014 0.441 1.769
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0702 0.1272 8.41 6.7e-16 ***
## FirstAuthorFemale1 0.0621 0.1122 0.55 0.58
## LastAuthorFemale1 -0.2074 0.1541 -1.35 0.18
## Year1997 0.0719 0.2057 0.35 0.73
## Year1998 0.1788 0.2071 0.86 0.39
## Year1999 0.1482 0.4178 0.35 0.72
## Year2000 0.1608 0.1688 0.95 0.34
## Year2001 0.3196 0.2078 1.54 0.12
## Year2002 -0.0438 0.1743 -0.25 0.80
## Year2003 0.1363 0.1650 0.83 0.41
## Year2004 -0.0150 0.1618 -0.09 0.93
## Year2005 0.1514 0.1720 0.88 0.38
```

```

## Year2006          -0.1653      0.1733   -0.95      0.34
## Year2007          -0.0775      0.1631   -0.47      0.64
## Year2008          -0.1109      0.1650   -0.67      0.50
## Year2009           0.0507      0.1645    0.31      0.76
## Year2010          -0.2318      0.1719   -1.35      0.18
## Year2011           0.1322      0.1765    0.75      0.45
## Year2012          -0.1953      0.1749   -1.12      0.26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.629
## Multiple R-squared:  0.0567, Adjusted R-squared:  0.0151
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 390 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.409  0.877  0.953  0.915  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.158 1      1.076
## Year              1.158 16      1.005

```

## Residuals from first author



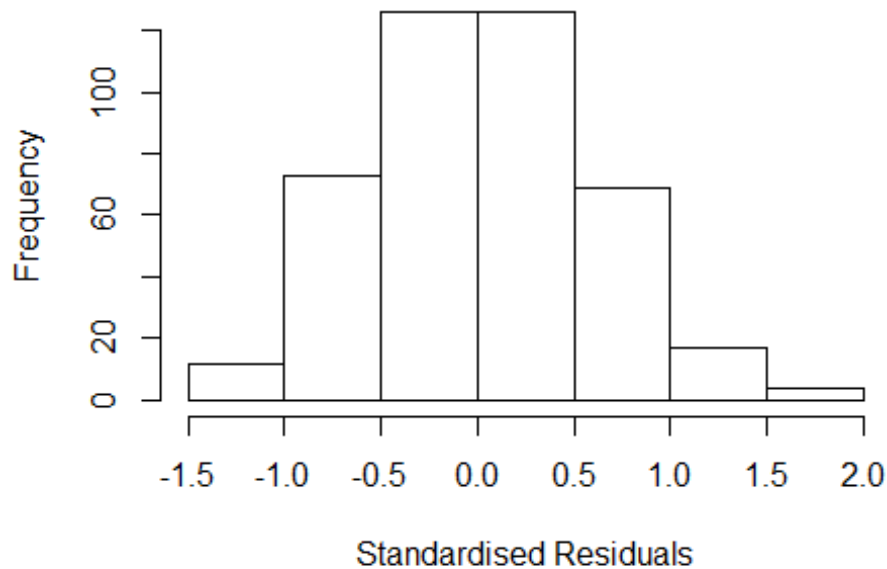
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2473 -0.4248 0.0113 0.4403 1.7707
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0701 0.1271 8.42 6.6e-16 ***
## FirstAuthorFemale1 -0.0311 0.1060 -0.29 0.77
## Year1997 0.0488 0.2025 0.24 0.81
## Year1998 0.1772 0.2108 0.84 0.40
## Year1999 0.1143 0.3927 0.29 0.77
## Year2000 0.1519 0.1701 0.89 0.37
## Year2001 0.3296 0.2089 1.58 0.12
## Year2002 -0.0604 0.1747 -0.35 0.73
## Year2003 0.1362 0.1650 0.83 0.41
## Year2004 -0.0128 0.1627 -0.08 0.94
## Year2005 0.1487 0.1701 0.87 0.38
## Year2006 -0.1629 0.1733 -0.94 0.35
```

```

## Year2007          -0.0747      0.1641   -0.46      0.65
## Year2008          -0.1326      0.1645   -0.81      0.42
## Year2009           0.0466      0.1658    0.28      0.78
## Year2010          -0.2314      0.1713   -1.35      0.18
## Year2011           0.1360      0.1770    0.77      0.44
## Year2012          -0.1934      0.1747   -1.11      0.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0524, Adjusted R-squared:  0.013
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 390 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.406  0.875   0.952   0.914   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.08 1          1.039
## Year              1.08 16          1.002

```

## Residuals from last author



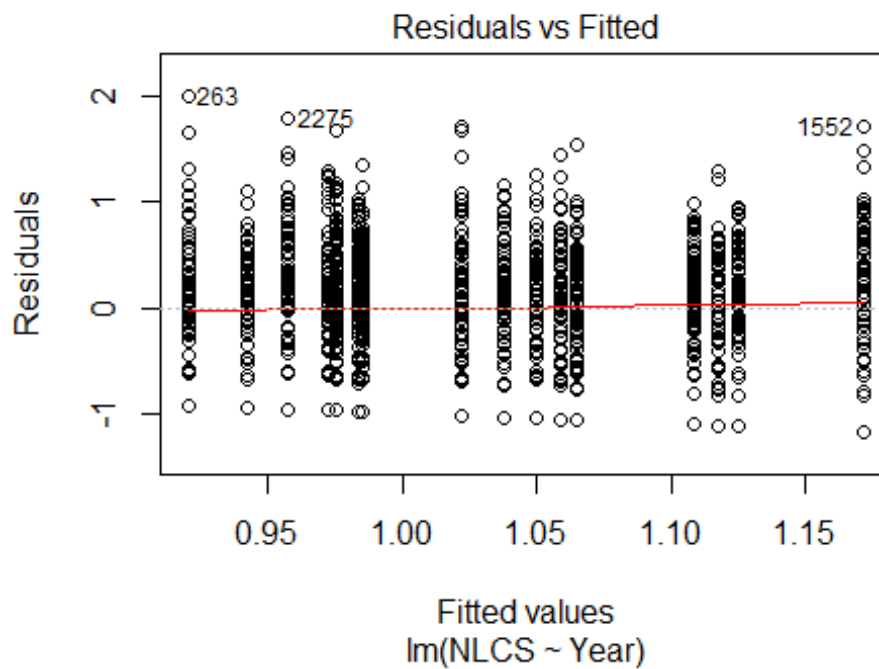
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2533 -0.4195 0.0122 0.4376 1.7647
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0701 0.1271 8.42 6.6e-16 ***
## LastAuthorFemale1 -0.1661 0.1332 -1.25 0.21
## Year1997 0.0726 0.2051 0.35 0.72
## Year1998 0.1832 0.2080 0.88 0.38
## Year1999 0.1410 0.4132 0.34 0.73
## Year2000 0.1627 0.1685 0.97 0.33
## Year2001 0.3262 0.2077 1.57 0.12
## Year2002 -0.0456 0.1737 -0.26 0.79
## Year2003 0.1362 0.1649 0.83 0.41
## Year2004 -0.0116 0.1616 -0.07 0.94
## Year2005 0.1531 0.1718 0.89 0.37
## Year2006 -0.1601 0.1732 -0.92 0.36
```

```

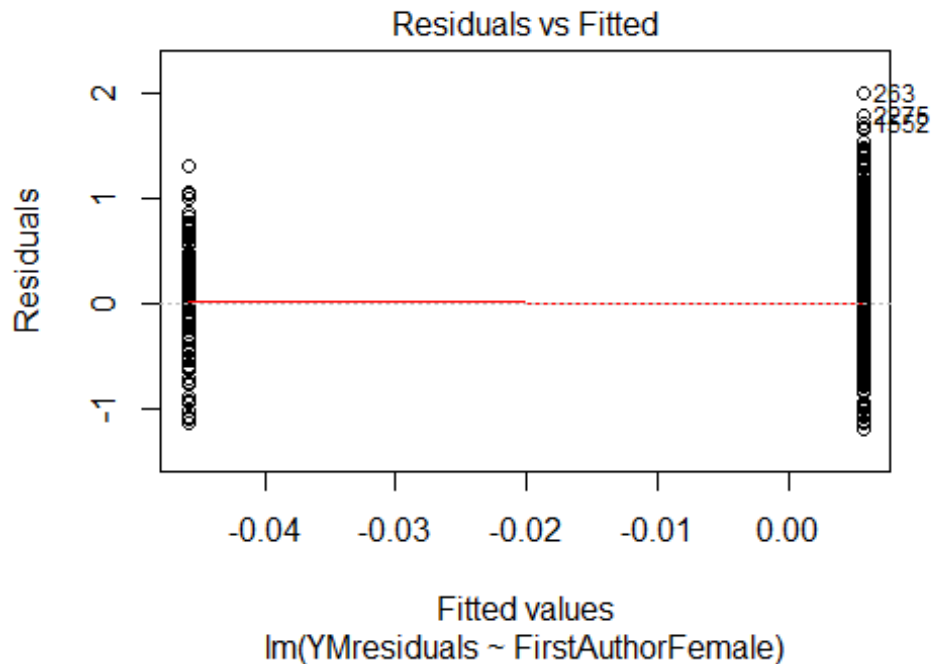
## Year2007          -0.0726      0.1628    -0.45      0.66
## Year2008          -0.1122      0.1651    -0.68      0.50
## Year2009           0.0583      0.1639     0.36      0.72
## Year2010          -0.2237      0.1705    -1.31      0.19
## Year2011           0.1391      0.1749     0.80      0.43
## Year2012          -0.1907      0.1747    -1.09      0.28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0561, Adjusted R-squared:  0.0168
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 35 weights are ~= 1. The remaining 392 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.409  0.876  0.952  0.915  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.34e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 427"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2610"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  217  180  223  228  266  254  245  327  253  194  209  228  217  184  134
## 2011 2012
##  166  193
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   87   55   83   73   97   98  102  117  104   82  109  115  112   89   62
## 2011 2012

```

```
##      80      85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   78   50   79   69   86   87   92  105   98   75  104  107   99   84   56
## 2011 2012
##   67   79
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 54, df = 16, p-value = 6e-06
```



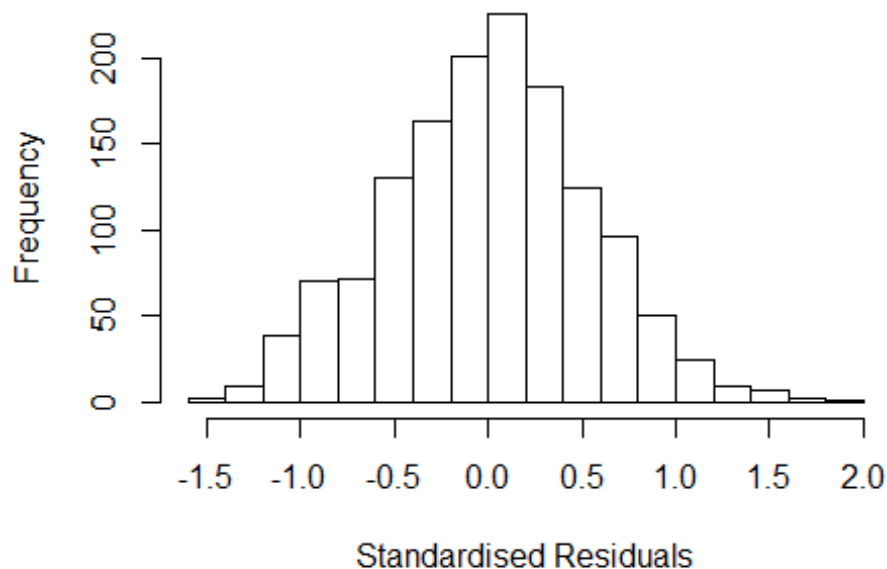
```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 1.81144732852781"
## [1] "Male first author team size 2018 geometric mean: 1.59407419678485"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 400, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.1689435423954"
## [1] "Male last author team size 2018 geometric mean: 1.58382702337824"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.391 1          1.180
## LastAuthorFemale  1.340 1          1.157
## UniqueAuthors     1.223 4          1.025
## Year               1.309 16         1.008
```



## Residuals from first and last author and team size



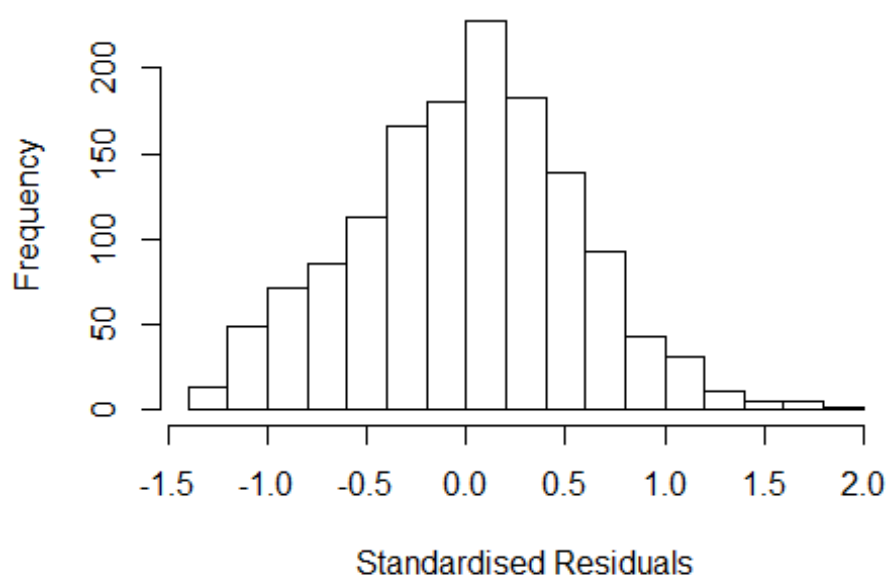
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4509 -0.3579 0.0158 0.3600 1.8126
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7858 0.0707 11.11 < 2e-16 ***
## FirstAuthorFemale1 -0.0559 0.0494 -1.13 0.2583
## LastAuthorFemale1 0.0580 0.0621 0.93 0.3506
## UniqueAuthors2 0.1812 0.0322 5.63 2.2e-08 ***
## UniqueAuthors3 0.3252 0.0576 5.65 2.0e-08 ***
## UniqueAuthors4 0.4461 0.1074 4.15 3.5e-05 ***
## UniqueAuthors5 0.3749 0.1834 2.04 0.0411 *
## Year1997 0.0653 0.1103 0.59 0.5540
## Year1998 0.0556 0.0887 0.63 0.5305
## Year1999 0.2287 0.0935 2.45 0.0146 *
```

```

## Year2000          0.1412      0.0898      1.57      0.1163
## Year2001          0.3400      0.1157      2.94      0.0034 **
## Year2002          0.2347      0.0837      2.81      0.0051 **
## Year2003          0.0089      0.1036      0.09      0.9316
## Year2004          0.1058      0.0963      1.10      0.2719
## Year2005          0.0397      0.0908      0.44      0.6623
## Year2006          0.1501      0.0892      1.68      0.0925 .
## Year2007          0.0604      0.0841      0.72      0.4733
## Year2008          0.0701      0.0839      0.84      0.4037
## Year2009          0.1560      0.0887      1.76      0.0790 .
## Year2010          0.2383      0.0968      2.46      0.0139 *
## Year2011          0.0877      0.0945      0.93      0.3533
## Year2012          0.1250      0.0911      1.37      0.1703
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.539
## Multiple R-squared:  0.075, Adjusted R-squared:  0.0604
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1284 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.235  0.865  0.951  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          7.07e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.377 1          1.173
## LastAuthorFemale 1.337 1          1.156
## Year              1.135 16          1.004

```

## Residuals from first and last author



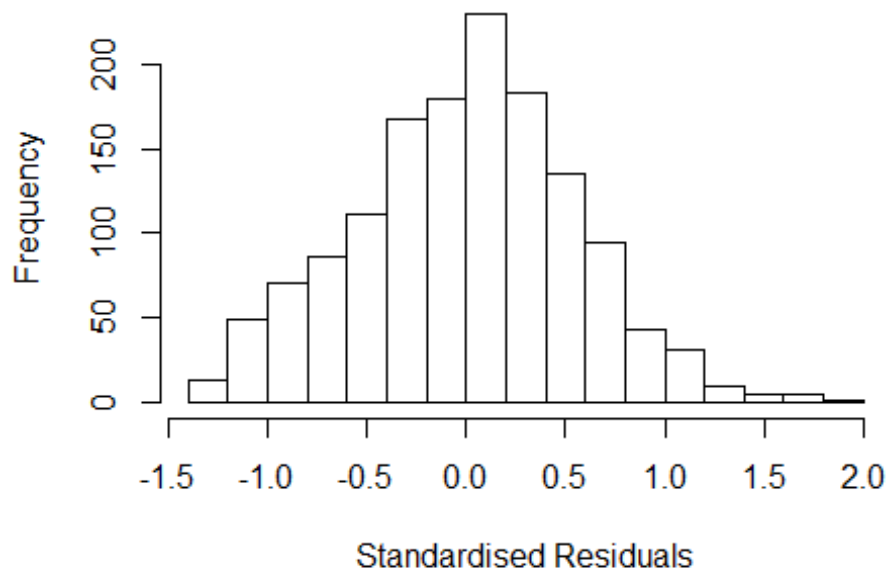
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2899 -0.3624  0.0236  0.3651  1.8318
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8759    0.0711   12.31 < 2e-16 ***
## FirstAuthorFemale1 -0.0109    0.0501   -0.22  0.82823
## LastAuthorFemale1  0.0207    0.0658    0.31  0.75312
## Year1997         0.0976    0.1169    0.84  0.40376
## Year1998         0.0793    0.0923    0.86  0.39072
## Year1999         0.2513    0.0950    2.65  0.00826 **
## Year2000         0.1604    0.0938    1.71  0.08756 .
## Year2001         0.3933    0.1156    3.40  0.00069 ***
## Year2002         0.2506    0.0873    2.87  0.00414 **
## Year2003         0.0512    0.1040    0.49  0.62238
## Year2004         0.1277    0.0979    1.30  0.19222
## Year2005         0.0329    0.0937    0.35  0.72572
```

```

## Year2006          0.1623      0.0917      1.77  0.07718 .
## Year2007          0.0869      0.0862      1.01  0.31350
## Year2008          0.0790      0.0871      0.91  0.36458
## Year2009          0.1874      0.0947      1.98  0.04804 *
## Year2010          0.2440      0.0998      2.44  0.01462 *
## Year2011          0.0842      0.0969      0.87  0.38520
## Year2012          0.1551      0.0937      1.65  0.09822 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.0154
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 123 weights are ~= 1. The remaining 1292 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.245  0.870  0.950  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.089 1      1.044
## Year              1.089 16      1.003

```

## Residuals from first author



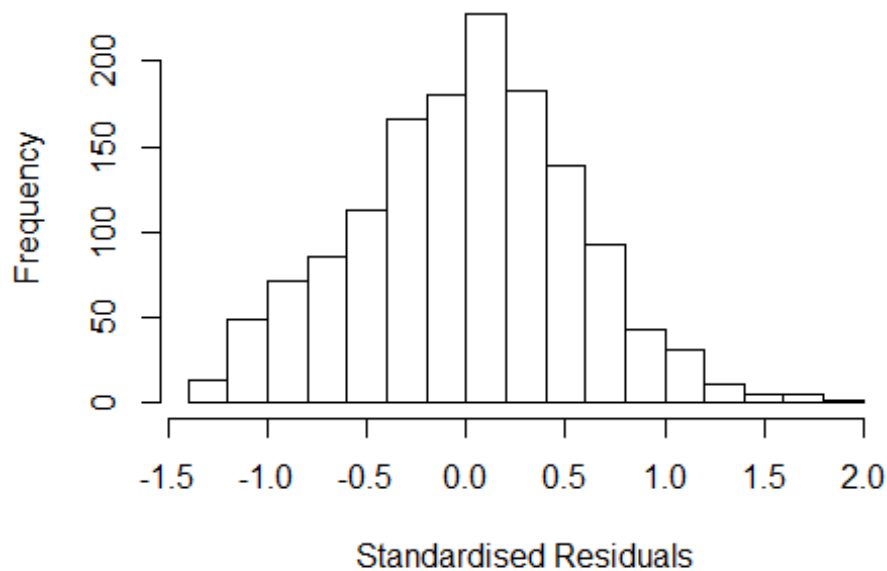
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2696 -0.3600  0.0229  0.3645  1.8320
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.87673    0.07115   12.32 < 2e-16 ***
## FirstAuthorFemale1 -0.00214    0.04462   -0.05  0.96166
## Year1997        0.09717    0.11667    0.83  0.40505
## Year1998        0.07951    0.09221    0.86  0.38872
## Year1999        0.25105    0.09493    2.64  0.00827 **
## Year2000        0.16025    0.09380    1.71  0.08780 .
## Year2001        0.39286    0.11549    3.40  0.00069 ***
## Year2002        0.25121    0.08706    2.89  0.00397 **
## Year2003        0.05031    0.10402    0.48  0.62871
## Year2004        0.12698    0.09788    1.30  0.19472
## Year2005        0.03212    0.09365    0.34  0.73170
## Year2006        0.16173    0.09172    1.76  0.07808 .
```

```

## Year2007          0.08783    0.08597    1.02  0.30715
## Year2008          0.07928    0.08689    0.91  0.36172
## Year2009          0.18739    0.09460    1.98  0.04779 *
## Year2010          0.24333    0.09978    2.44  0.01487 *
## Year2011          0.08366    0.09689    0.86  0.38800
## Year2012          0.15563    0.09368    1.66  0.09688 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0278, Adjusted R-squared:  0.016
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1291 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.244  0.869  0.950  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.056 1          1.027
## Year            1.056 16          1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2827 -0.3614  0.0241  0.3658  1.8330
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8758    0.0712   12.29 < 2e-16 ***
## LastAuthorFemale1 0.0145    0.0586    0.25  0.80499
## Year1997        0.0972    0.1168    0.83  0.40555
## Year1998        0.0792    0.0924    0.86  0.39153
## Year1999        0.2497    0.0945    2.64  0.00831 **
## Year2000        0.1591    0.0934    1.70  0.08854 .
## Year2001        0.3925    0.1154    3.40  0.00069 ***
## Year2002        0.2506    0.0874    2.87  0.00418 **
## Year2003        0.0502    0.1040    0.48  0.62900
## Year2004        0.1274    0.0979    1.30  0.19328
## Year2005        0.0317    0.0933    0.34  0.73456
## Year2006        0.1611    0.0913    1.77  0.07762 .
```

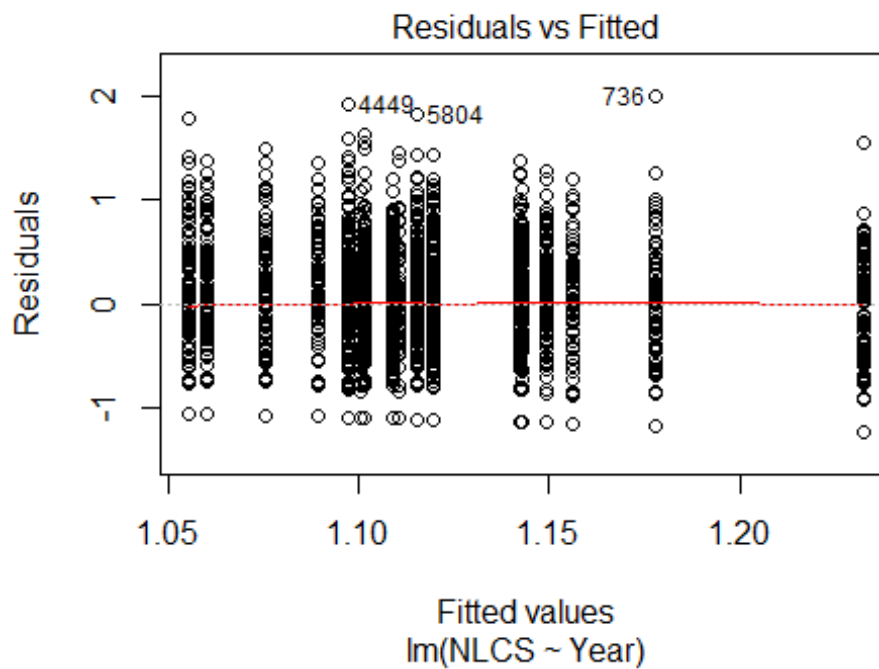
```

## Year2007          0.0867      0.0863      1.00  0.31524
## Year2008          0.0787      0.0871      0.90  0.36629
## Year2009          0.1871      0.0947      1.97  0.04848 *
## Year2010          0.2435      0.0998      2.44  0.01483 *
## Year2011          0.0832      0.0966      0.86  0.38935
## Year2012          0.1550      0.0938      1.65  0.09863 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0279, Adjusted R-squared:  0.0161
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1291 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.244  0.869  0.950  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1415"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2611"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 337 306 267 268 299 291 235 250 246 212 283 385 419 480 470
## 2011 2012
## 455 480
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 95 80 77 78 100 89 77 87 92 125 176 205 226 227
## 2011 2012

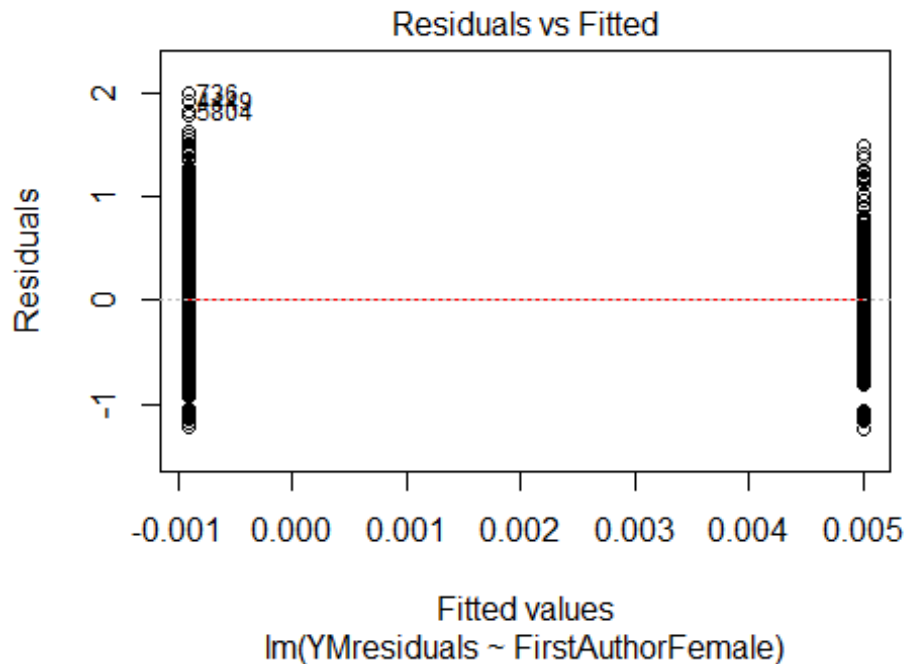
```



```
## 247 278
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 88 85 75 72 68 91 74 68 77 78 103 152 168 199 191
## 2011 2012
## 209 237
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```

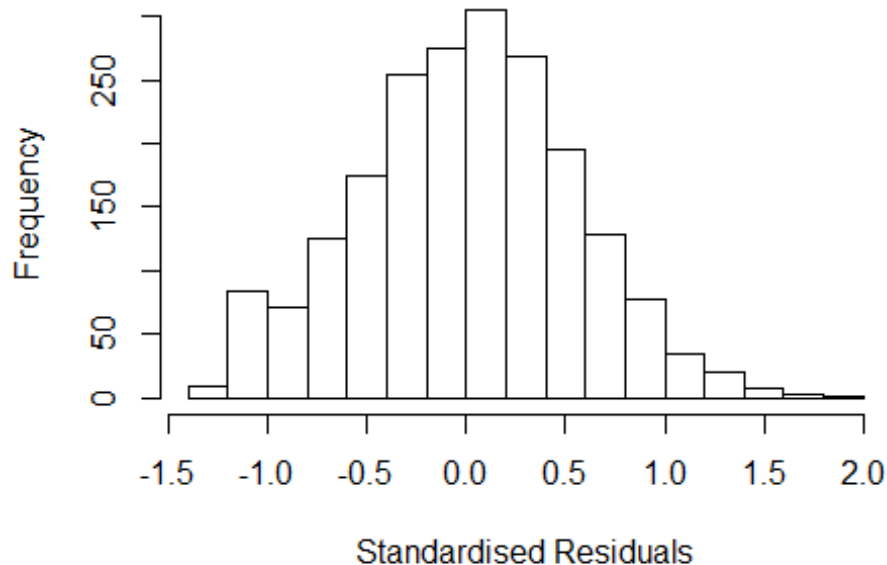


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.86658099007102"
## [1] "Male first author team size 2018 geometric mean: 2.48533529716968"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6500, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.6280448956487"
## [1] "Male last author team size 2018 geometric mean: 2.54153735411315"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4900, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.085 1      1.042
## LastAuthorFemale  1.093 1      1.046
## UniqueAuthors    1.234 4      1.027
## Year             1.264 16      1.007
```

## Residuals from first and last author and team size



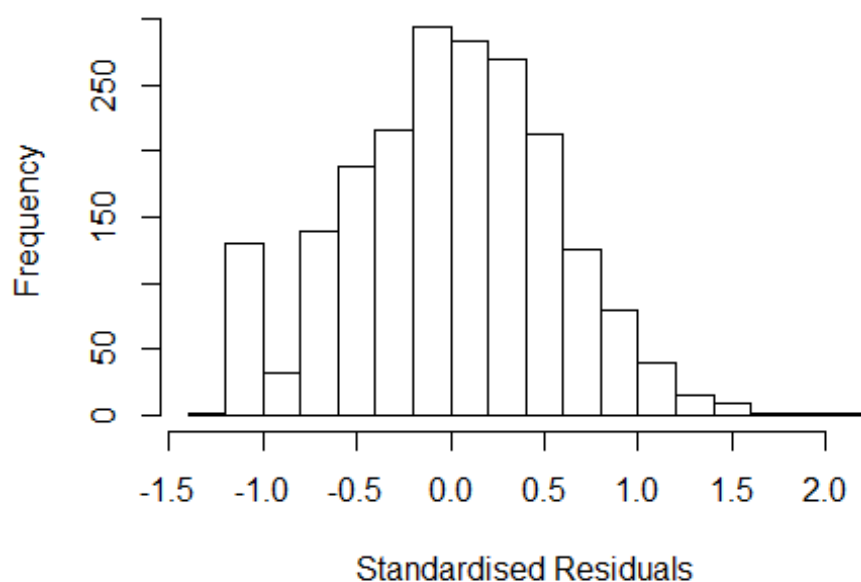
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2952 -0.3615 0.0128 0.3692 1.9617
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00457 0.07841 12.81 < 2e-16 ***
## FirstAuthorFemale1 -0.02206 0.03399 -0.65 0.52
## LastAuthorFemale1 0.04004 0.03833 1.04 0.30
## UniqueAuthors2 0.15740 0.03338 4.72 2.6e-06 ***
## UniqueAuthors3 0.16179 0.03700 4.37 1.3e-05 ***
## UniqueAuthors4 0.31456 0.04982 6.31 3.3e-10 ***
## UniqueAuthors5 0.37826 0.04722 8.01 1.9e-15 ***
## Year1997 0.01592 0.11433 0.14 0.89
## Year1998 0.05831 0.10753 0.54 0.59
## Year1999 0.07832 0.10378 0.75 0.45
```

```

## Year2000      0.04031    0.10389    0.39    0.70
## Year2001     -0.10436    0.10517   -0.99    0.32
## Year2002      0.17010    0.09708    1.75    0.08 .
## Year2003     -0.05420    0.09486   -0.57    0.57
## Year2004     -0.04201    0.10050   -0.42    0.68
## Year2005     -0.08400    0.11522   -0.73    0.47
## Year2006     -0.00596    0.09312   -0.06    0.95
## Year2007     -0.02819    0.09036   -0.31    0.76
## Year2008     -0.07883    0.09005   -0.88    0.38
## Year2009      0.01135    0.08527    0.13    0.89
## Year2010     -0.06577    0.08784   -0.75    0.45
## Year2011     -0.08758    0.08643   -1.01    0.31
## Year2012     -0.06280    0.08556   -0.73    0.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0446, Adjusted R-squared:  0.0341
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 170 weights are ~= 1. The remaining 1865 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.166  0.867  0.951  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061 1      1.030
## LastAuthorFemale  1.077 1      1.038
## Year              1.070 16      1.002

```

## Residuals from first and last author



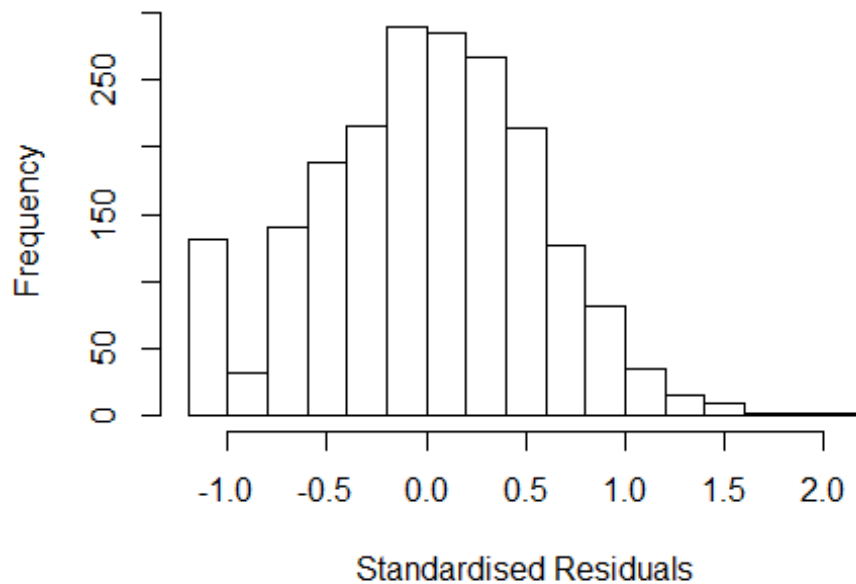
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2305 -0.3824 0.0164 0.3820 2.0416
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07099 0.08079 13.26 <2e-16 ***
## FirstAuthorFemale1 0.00890 0.03476 0.26 0.798
## LastAuthorFemale1 0.04574 0.04031 1.13 0.257
## Year1997 0.03776 0.11821 0.32 0.749
## Year1998 0.06942 0.11024 0.63 0.529
## Year1999 0.10488 0.10540 1.00 0.320
## Year2000 0.08729 0.10417 0.84 0.402
## Year2001 -0.05955 0.10750 -0.55 0.580
## Year2002 0.20606 0.09928 2.08 0.038 *
## Year2003 0.00901 0.09721 0.09 0.926
## Year2004 0.02898 0.10326 0.28 0.779
## Year2005 -0.01417 0.11854 -0.12 0.905
```

```

## Year2006          0.06292      0.09615      0.65      0.513
## Year2007          0.03862      0.09313      0.41      0.678
## Year2008         -0.03113      0.09267     -0.34      0.737
## Year2009          0.08137      0.08816      0.92      0.356
## Year2010          0.02311      0.08957      0.26      0.796
## Year2011         -0.00756      0.08842     -0.09      0.932
## Year2012          0.03117      0.08774      0.36      0.722
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.00945,    Adjusted R-squared:  0.000607
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 163 weights are ~= 1. The remaining 1872 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.159  0.865   0.951   0.909   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year              1.024 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1995 -0.3783 0.0196 0.3782 2.0367
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07772 0.08011 13.45 <2e-16 ***
## FirstAuthorFemale1 0.01869 0.03398 0.55 0.582
## Year1997 0.03607 0.11804 0.31 0.760
## Year1998 0.06759 0.11019 0.61 0.540
## Year1999 0.10311 0.10556 0.98 0.329
## Year2000 0.08622 0.10392 0.83 0.407
## Year2001 -0.06438 0.10724 -0.60 0.548
## Year2002 0.20023 0.09875 2.03 0.043 *
## Year2003 0.00472 0.09682 0.05 0.961
## Year2004 0.02399 0.10294 0.23 0.816
## Year2005 -0.01427 0.11824 -0.12 0.904
## Year2006 0.05946 0.09608 0.62 0.536
```

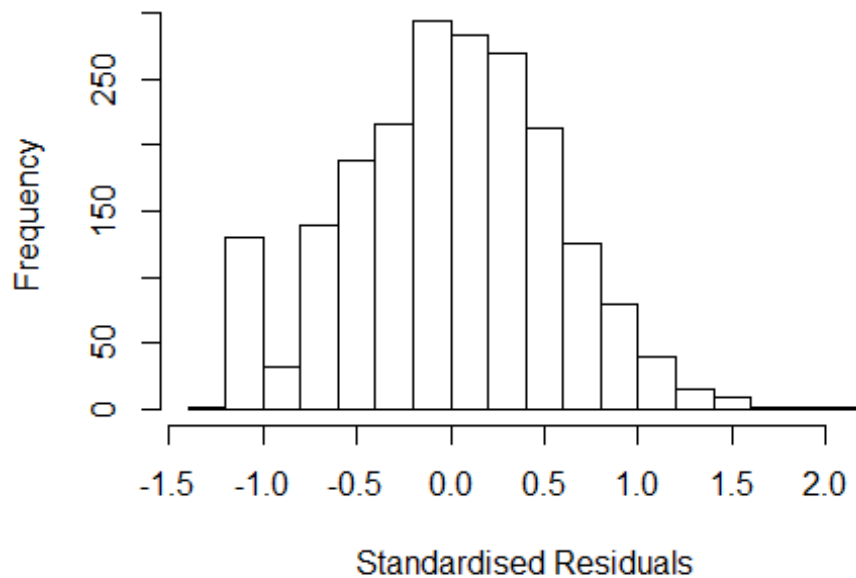
```

## Year2007          0.03483    0.09279    0.38    0.707
## Year2008          -0.03718    0.09203   -0.40    0.686
## Year2009          0.07984    0.08811    0.91    0.365
## Year2010          0.01971    0.08937    0.22    0.825
## Year2011          -0.01062    0.08813   -0.12    0.904
## Year2012          0.02945    0.08768    0.34    0.737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.00885,    Adjusted R-squared:  0.000497
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 165 weights are ~= 1. The remaining 1870 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.161  0.865  0.950  0.909  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.04 1      1.020
## Year              1.04 16      1.001

```



## Residuals from last author



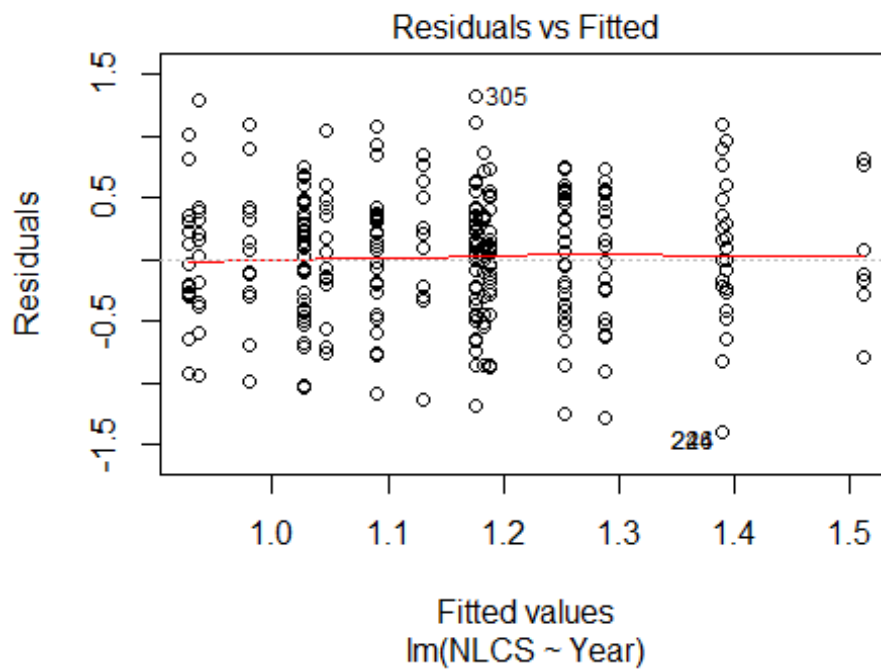
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2253 -0.3816 0.0162 0.3815 2.0413
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07169 0.08079 13.26 <2e-16 ***
## LastAuthorFemale1 0.04821 0.03960 1.22 0.224
## Year1997 0.03768 0.11818 0.32 0.750
## Year1998 0.06905 0.11025 0.63 0.531
## Year1999 0.10538 0.10535 1.00 0.317
## Year2000 0.08767 0.10420 0.84 0.400
## Year2001 -0.05924 0.10744 -0.55 0.581
## Year2002 0.20631 0.09926 2.08 0.038 *
## Year2003 0.00983 0.09709 0.10 0.919
## Year2004 0.02937 0.10317 0.28 0.776
## Year2005 -0.01437 0.11845 -0.12 0.903
## Year2006 0.06360 0.09608 0.66 0.508
```

```

## Year2007      0.03895    0.09312    0.42    0.676
## Year2008     -0.03066    0.09258   -0.33    0.741
## Year2009      0.08201    0.08805    0.93    0.352
## Year2010      0.02344    0.08949    0.26    0.793
## Year2011     -0.00676    0.08829   -0.08    0.939
## Year2012      0.03169    0.08765    0.36    0.718
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.00942,    Adjusted R-squared:  0.00107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1874 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.159  0.866  0.952  0.909  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2035"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2612"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   65   58   55   52   54   42   53   37   47   42   58   59   48   59   49
## 2011 2012
##   65   60
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   13   12   14   13    6    8   14   13   15   15   27   13   23   31   21
## 2011 2012

```

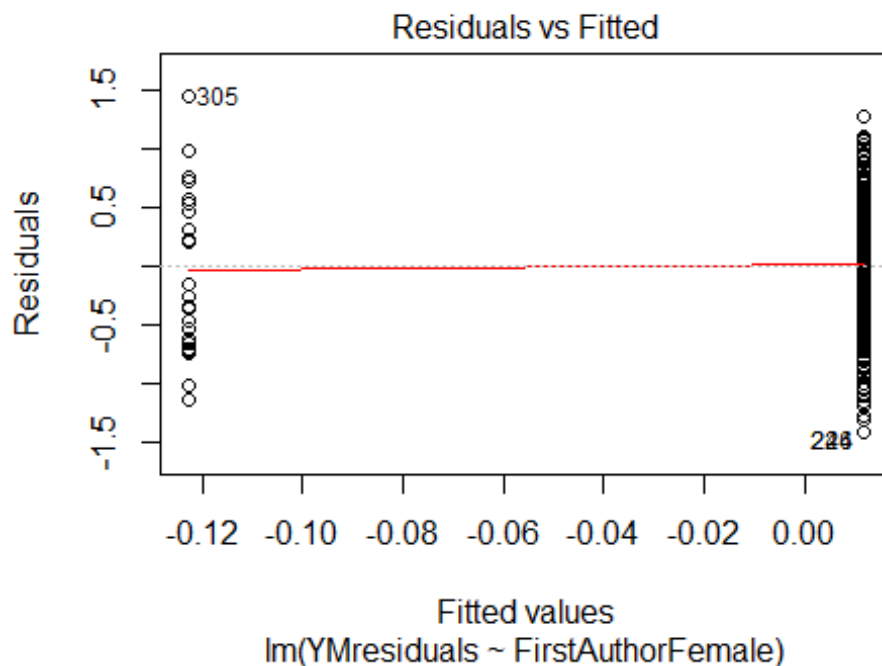
```
## 29 30
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 10 10 13 6 7 14 13 14 14 23 13 22 26 21
## 2011 2012
## 22 28
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
## [1] "Female first author team size 2018 geometric mean: 1.41421356237309"
## [1] "Male first author team size 2018 geometric mean: 2.02169923078425"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

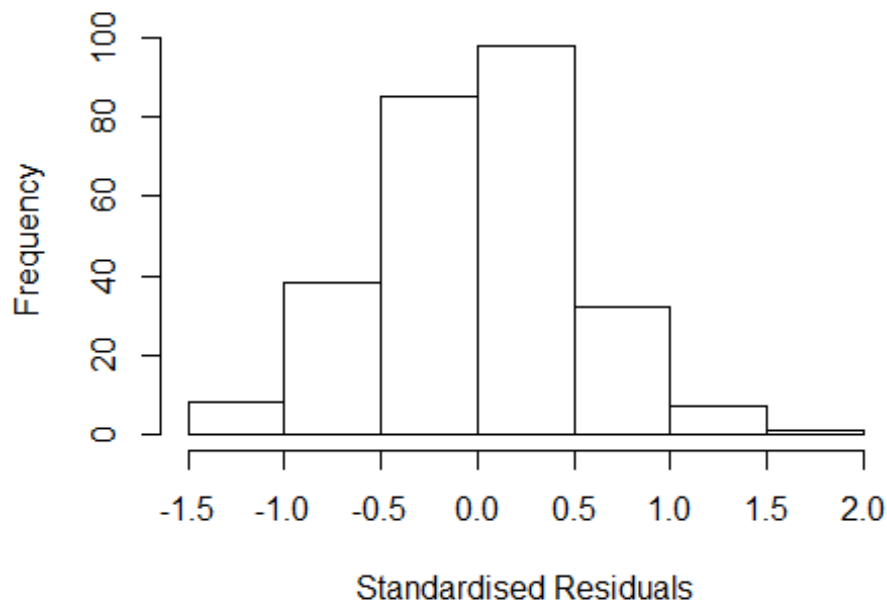
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.25992104989487"
## [1] "Male last author team size 2018 geometric mean: 2.02048756505036"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.726 1      1.314
## LastAuthorFemale  1.487 1      1.219
## UniqueAuthors    2.279 4      1.108
## Year              3.208 16     1.037
```

## Residuals from first and last author and team size



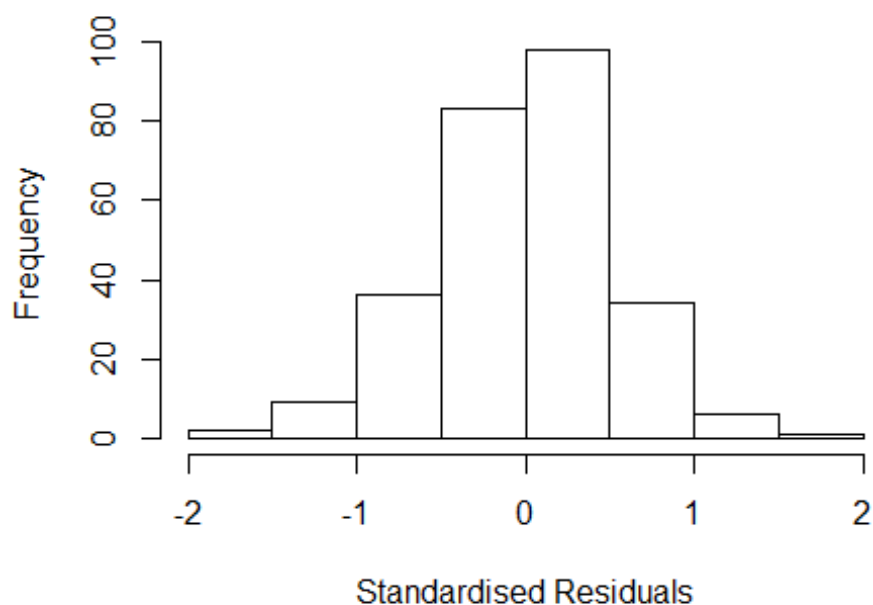
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4624 -0.3360 0.0213 0.3548 1.6735
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1509 0.1747 6.59 2.7e-10 ***
## FirstAuthorFemale1 -0.0269 0.1921 -0.14 0.8886
## LastAuthorFemale1 -0.1818 0.1709 -1.06 0.2885
## UniqueAuthors2 0.1223 0.0693 1.76 0.0788 .
## UniqueAuthors3 0.1124 0.1343 0.84 0.4037
## UniqueAuthors4 0.8001 0.2865 2.79 0.0056 **
## UniqueAuthors5 0.2098 0.2075 1.01 0.3128
## Year1997 0.0669 0.2162 0.31 0.7572
## Year1998 0.1039 0.2640 0.39 0.6943
## Year1999 0.3115 0.2820 1.10 0.2705
```

```

## Year2000          -0.1177      0.3767   -0.31   0.7549
## Year2001           0.3726      0.3015    1.24   0.2177
## Year2002          -0.2762      0.2259   -1.22   0.2227
## Year2003          -0.3084      0.2429   -1.27   0.2055
## Year2004          -0.2012      0.2281   -0.88   0.3787
## Year2005          -0.2197      0.2326   -0.94   0.3460
## Year2006          -0.1750      0.1948   -0.90   0.3698
## Year2007          -0.1647      0.2290   -0.72   0.4728
## Year2008           0.0154      0.1981    0.08   0.9383
## Year2009          -0.1635      0.2005   -0.82   0.4155
## Year2010           0.0859      0.2245    0.38   0.7022
## Year2011          -0.0674      0.2075   -0.32   0.7457
## Year2012           0.1031      0.2019    0.51   0.6101
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.123, Adjusted R-squared:  0.0445
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 252 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.284  0.884  0.956  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.577 1      1.256
## LastAuthorFemale  1.451 1      1.205
## Year              1.504 16      1.013

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5287 -0.3232 0.0169 0.3685 1.6244
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18e+00 1.79e-01 6.61 2.4e-10 ***
## FirstAuthorFemale1 -1.99e-02 1.88e-01 -0.11 0.92
## LastAuthorFemale1 -1.88e-01 1.66e-01 -1.14 0.26
## Year1997 1.05e-01 2.22e-01 0.47 0.64
## Year1998 1.35e-01 2.74e-01 0.49 0.62
## Year1999 3.47e-01 2.89e-01 1.20 0.23
## Year2000 -9.97e-02 3.52e-01 -0.28 0.78
## Year2001 3.99e-01 3.02e-01 1.32 0.19
## Year2002 -2.39e-01 2.32e-01 -1.03 0.31
## Year2003 -2.55e-01 2.36e-01 -1.08 0.28
## Year2004 -1.83e-01 2.31e-01 -0.79 0.43
## Year2005 -1.86e-01 2.51e-01 -0.74 0.46
```

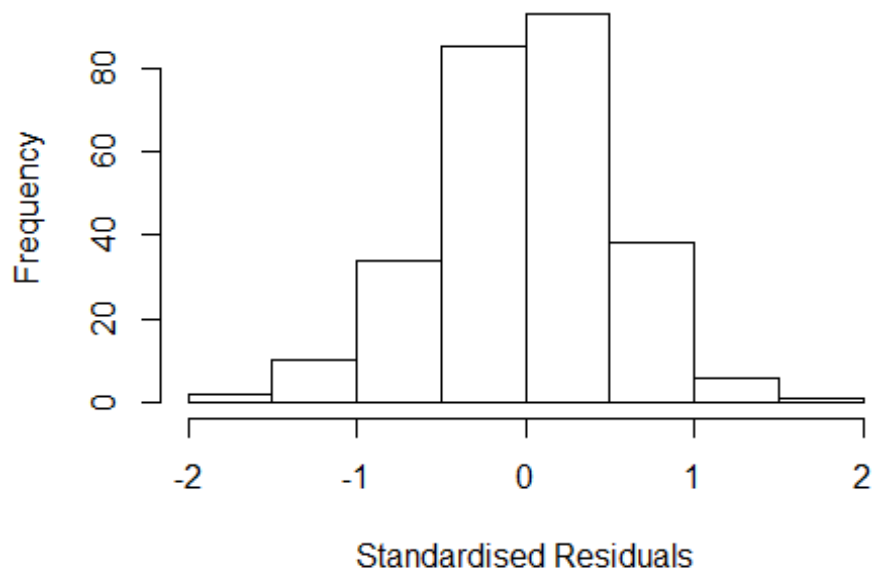
```

## Year2006      -1.52e-01   1.99e-01   -0.76    0.45
## Year2007      -1.13e-01   2.35e-01   -0.48    0.63
## Year2008       4.40e-02   2.01e-01    0.22    0.83
## Year2009      -1.47e-01   2.08e-01   -0.71    0.48
## Year2010       1.52e-01   2.28e-01    0.67    0.50
## Year2011      -6.15e-05   2.10e-01    0.00    1.00
## Year2012       1.29e-01   2.05e-01    0.63    0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.101, Adjusted R-squared:  0.036
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 255 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.309  0.872  0.957  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.269 1      1.126
## Year              1.269 16      1.007

```



## Residuals from first author



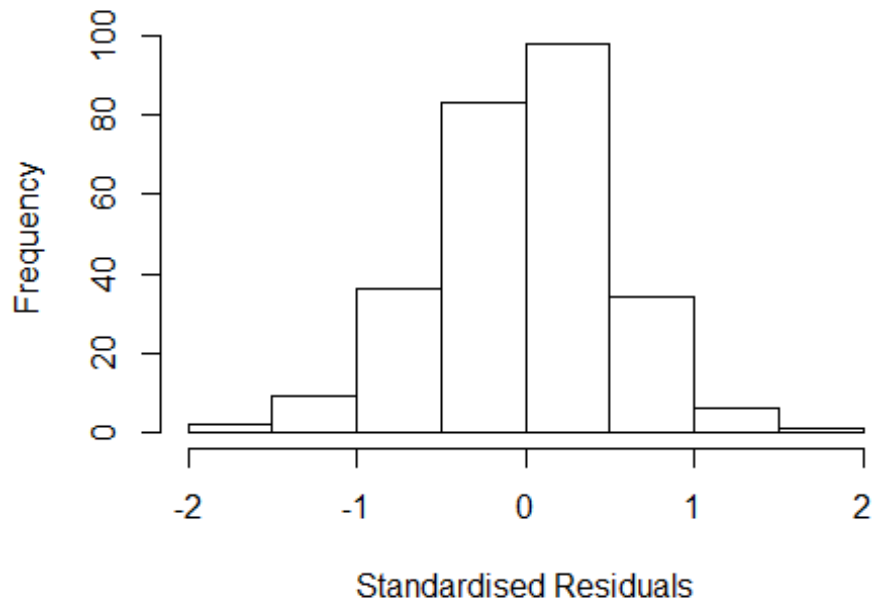
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5129 -0.3306 0.0301 0.3612 1.5037
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17717 0.18162 6.48 4.8e-10 ***
## FirstAuthorFemale1 -0.10556 0.17286 -0.61 0.54
## Year1997 0.13281 0.22896 0.58 0.56
## Year1998 0.12118 0.26834 0.45 0.65
## Year1999 0.33573 0.29521 1.14 0.26
## Year2000 -0.07735 0.36018 -0.21 0.83
## Year2001 0.38448 0.29770 1.29 0.20
## Year2002 -0.26441 0.22908 -1.15 0.25
## Year2003 -0.24539 0.23971 -1.02 0.31
## Year2004 -0.17290 0.23537 -0.73 0.46
## Year2005 -0.18286 0.25199 -0.73 0.47
## Year2006 -0.16484 0.20201 -0.82 0.42
```

```

## Year2007          -0.12699    0.23856   -0.53    0.59
## Year2008          0.03220    0.20272    0.16    0.87
## Year2009         -0.14438    0.20971   -0.69    0.49
## Year2010          0.15250    0.22782    0.67    0.50
## Year2011          0.00376    0.21495    0.02    0.99
## Year2012          0.11603    0.20779    0.56    0.58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.0909, Adjusted R-squared:  0.0293
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 254 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.407  0.883  0.957  0.911  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.186 1          1.089
## Year            1.186 16          1.005

```

## Residuals from last author



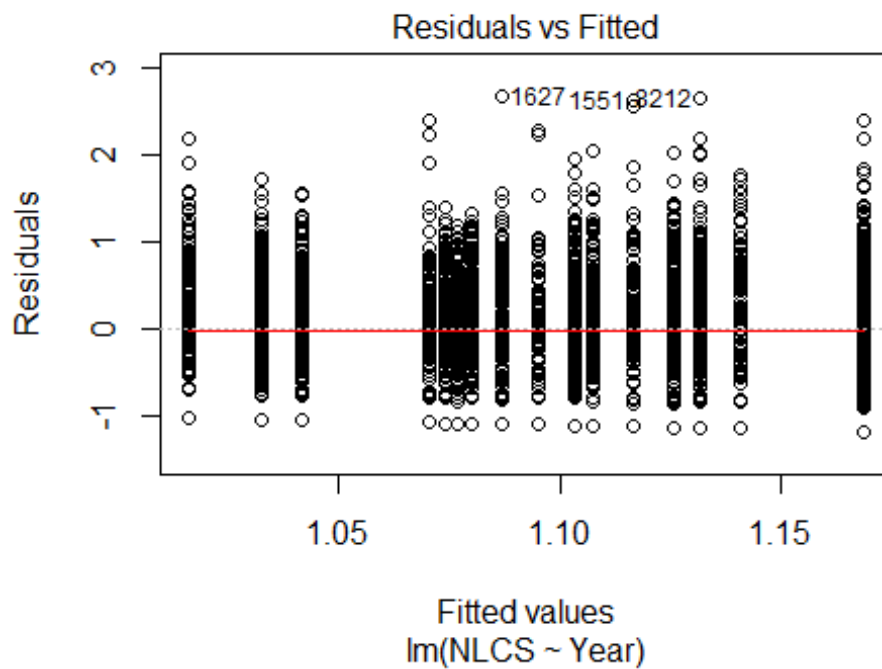
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5208 -0.3196 0.0242 0.3689 1.6042
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17919 0.17840 6.61 2.3e-10 ***
## LastAuthorFemale1 -0.19281 0.15144 -1.27 0.20
## Year1997 0.10174 0.21416 0.48 0.64
## Year1998 0.13381 0.27177 0.49 0.62
## Year1999 0.34163 0.28898 1.18 0.24
## Year2000 -0.09261 0.36068 -0.26 0.80
## Year2001 0.39862 0.29761 1.34 0.18
## Year2002 -0.23450 0.23284 -1.01 0.31
## Year2003 -0.25588 0.22972 -1.11 0.27
## Year2004 -0.18084 0.23027 -0.79 0.43
## Year2005 -0.18473 0.24998 -0.74 0.46
## Year2006 -0.14965 0.19922 -0.75 0.45
```

```

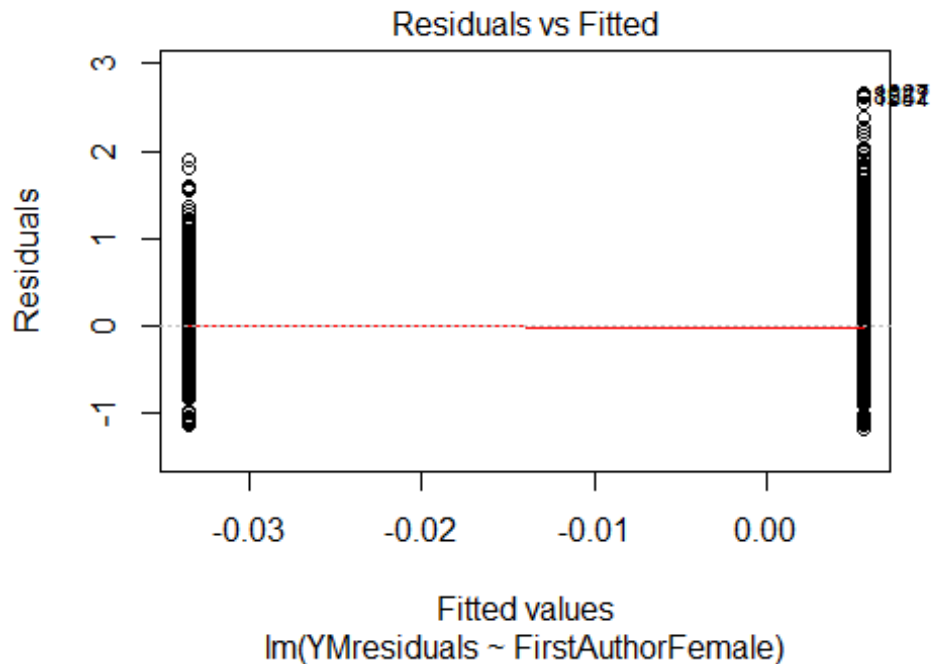
## Year2007      -0.11211      0.23431      -0.48      0.63
## Year2008      0.04496      0.20100      0.22      0.82
## Year2009     -0.14571      0.20693     -0.70      0.48
## Year2010      0.15151      0.22627      0.67      0.50
## Year2011     -0.00155      0.20848     -0.01      0.99
## Year2012      0.12881      0.20532      0.63      0.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0983, Adjusted R-squared:  0.0372
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 254 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.343  0.875  0.960  0.910  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.72e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 269"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2613"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  293  249  264  213  258  340  369  306  302  406  298  529  590  594  534
## 2011 2012
##  574  608
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  178  128  149  108  133  172  224  195  194  277  177  309  367  348  323
## 2011 2012

```

```
## 359 396
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 169 121 139 103 121 155 198 179 181 253 158 274 333 311 296
## 2011 2012
## 329 355
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 88, df = 16, p-value = 6e-12
```

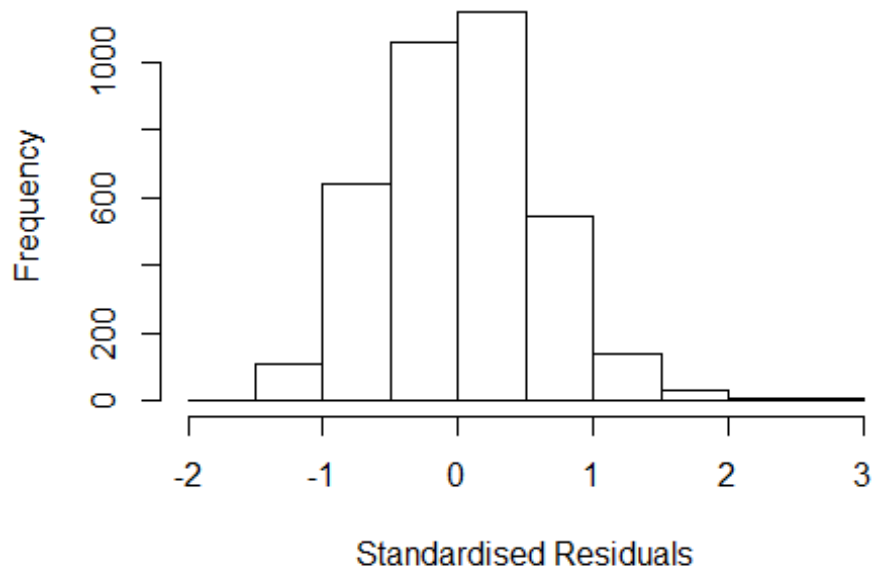


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.41417320295996"
## [1] "Male first author team size 2018 geometric mean: 2.06945978373048"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12000, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.36876278745065"
## [1] "Male last author team size 2018 geometric mean: 2.08109167450273"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.158 1      1.076
## LastAuthorFemale  1.164 1      1.079
## UniqueAuthors     1.119 4      1.014
## Year              1.126 16     1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 371   0031466983 3.453 1997    1100     5      2.525
## 1627  0001224048 3.749 2001    1702     4      2.873
## 7445  77749249761 3.553 2010    1712     3      2.617
## 8212  78650862532 3.769 2011    1804     2      2.887
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5342 -0.3886  0.0181  0.3869  2.8871
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.86994    0.06322   13.76  <2e-16 ***
## FirstAuthorFemale1 -0.07660    0.02987   -2.56    0.01 *
## LastAuthorFemale1 -0.02243    0.03356   -0.67    0.50
## UniqueAuthors2     0.28573    0.02402   11.90  <2e-16 ***
## UniqueAuthors3     0.35262    0.02796   12.61  <2e-16 ***
## UniqueAuthors4     0.51456    0.04346   11.84  <2e-16 ***
## UniqueAuthors5     0.60089    0.04927   12.20  <2e-16 ***
```

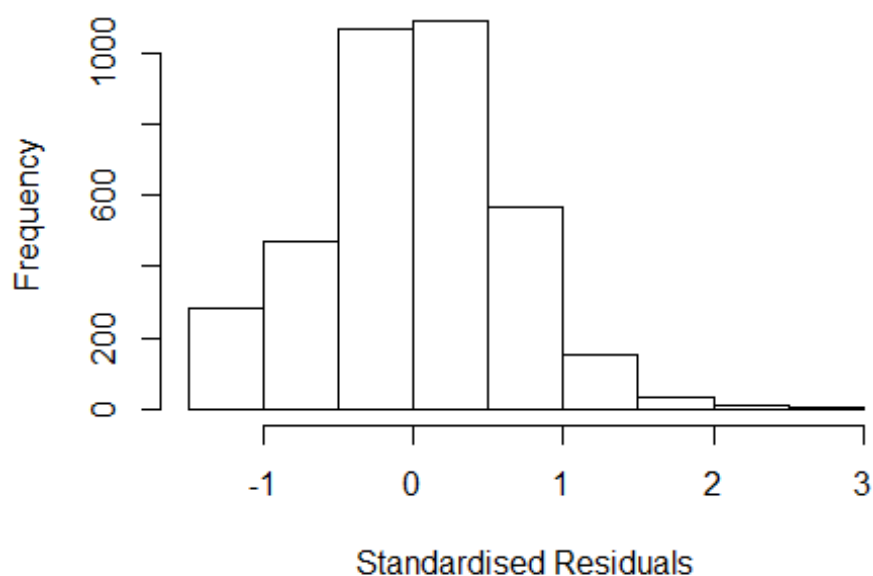
```

## Year1997      0.05762    0.08288    0.70    0.49
## Year1998      0.14969    0.09306    1.61    0.11
## Year1999      0.03525    0.08968    0.39    0.69
## Year2000      0.03016    0.09156    0.33    0.74
## Year2001      0.00653    0.08079    0.08    0.94
## Year2002      0.09290    0.07447    1.25    0.21
## Year2003      0.03427    0.07299    0.47    0.64
## Year2004      0.00528    0.07724    0.07    0.95
## Year2005     -0.04101    0.07284   -0.56    0.57
## Year2006      0.00780    0.07534    0.10    0.92
## Year2007      0.03417    0.07140    0.48    0.63
## Year2008     -0.01315    0.06972   -0.19    0.85
## Year2009      0.02038    0.07139    0.29    0.78
## Year2010      0.06599    0.07331    0.90    0.37
## Year2011      0.01194    0.07116    0.17    0.87
## Year2012     -0.02565    0.07135   -0.36    0.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.585
## Multiple R-squared:  0.096, Adjusted R-squared:  0.0905
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(661,3289) are outliers with |weight| = 0 ( < 2.7e-05);
## 299 weights are ~ = 1. The remaining 3374 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0076 0.8620 0.9510 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.110 1      1.054
## LastAuthorFemale 1.126 1      1.061
## Year 1.040 16      1.001

```



## Residuals from first and last author



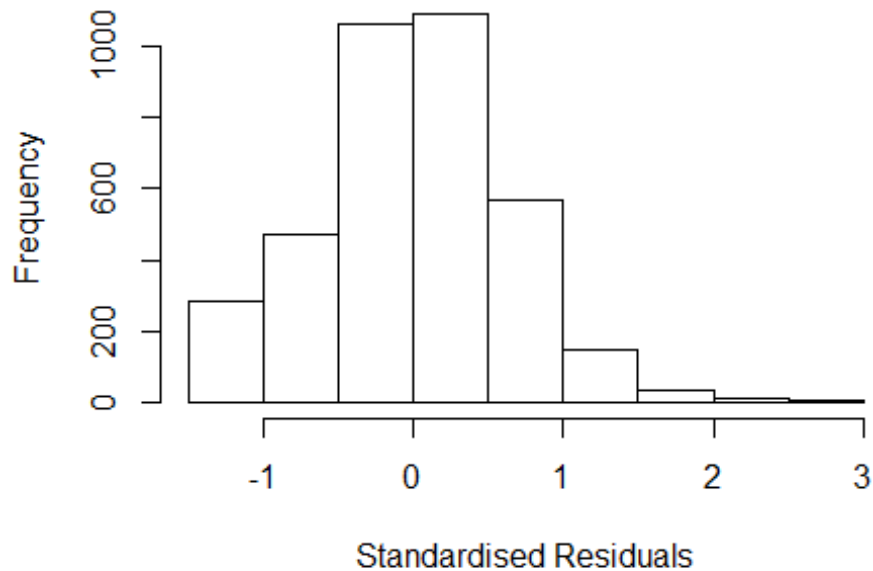
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1425  0001460136 3.678 2000    1703     4    2.611
## 1551  0006407254 3.743 2000    1703     4    2.676
## 1627  0001224048 3.749 2001    1702     4    2.710
## 8212  78650862532 3.769 2011    1804     2    2.660
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.16961 -0.41043  0.00624  0.41440  2.70972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9699    0.0654   14.84  <2e-16 ***
## FirstAuthorFemale1 -0.0141    0.0309   -0.45  0.6491
## LastAuthorFemale1 -0.0127    0.0358   -0.36  0.7224
## Year1997         0.0609    0.0877    0.69  0.4877
## Year1998         0.1497    0.0965    1.55  0.1208
## Year1999         0.0863    0.0936    0.92  0.3568
## Year2000         0.0967    0.0966    1.00  0.3166
## Year2001         0.0694    0.0839    0.83  0.4082
## Year2002         0.1897    0.0766    2.48  0.0133 *
```

```

## Year2003          0.1086      0.0754      1.44      0.1500
## Year2004          0.1074      0.0804      1.34      0.1814
## Year2005          0.0550      0.0757      0.73      0.4673
## Year2006          0.0895      0.0791      1.13      0.2578
## Year2007          0.1208      0.0731      1.65      0.0984 .
## Year2008          0.0733      0.0721      1.02      0.3093
## Year2009          0.1322      0.0744      1.78      0.0759 .
## Year2010          0.1997      0.0754      2.65      0.0081 **
## Year2011          0.1390      0.0738      1.88      0.0598 .
## Year2012          0.1068      0.0744      1.44      0.1508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.00604,    Adjusted R-squared:  0.00114
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 324 weights are ~= 1. The remaining 3351 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0119 0.8690 0.9490 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1          1.006
## Year              1.012 16          1.000

```

## Residuals from first author



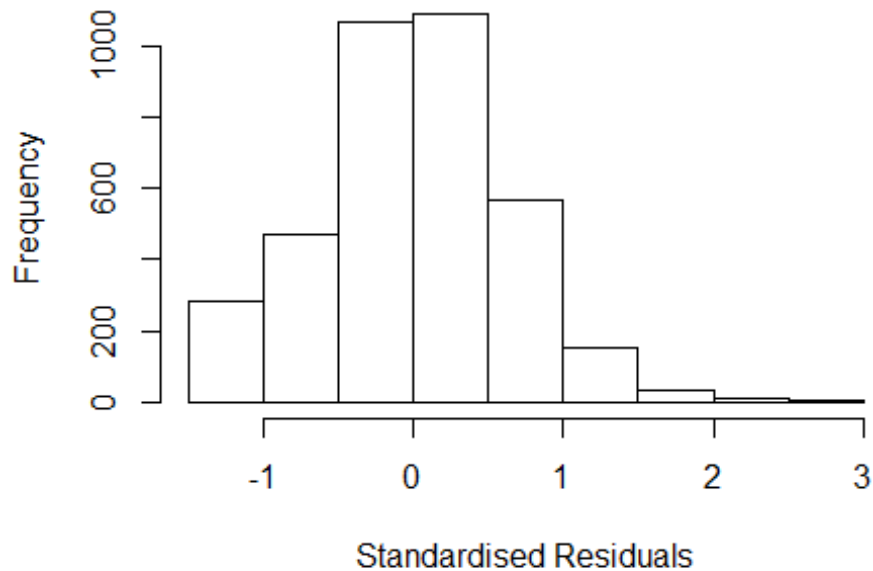
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1425  0001460136 3.678 2000    1703     4    2.611
## 1551  0006407254 3.743 2000    1703     4    2.676
## 1627  0001224048 3.749 2001    1702     4    2.710
## 8212  78650862532 3.769 2011    1804     2    2.660
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1686 -0.4096  0.0076  0.4166  2.7103
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9685    0.0651   14.87  <2e-16 ***
## FirstAuthorFemale1 -0.0178    0.0296   -0.60    0.547
## Year1997         0.0601    0.0878    0.69    0.493
## Year1998         0.1495    0.0965    1.55    0.121
## Year1999         0.0868    0.0936    0.93    0.354
## Year2000         0.0972    0.0966    1.01    0.315
## Year2001         0.0702    0.0838    0.84    0.402
## Year2002         0.1902    0.0765    2.49    0.013 *
## Year2003         0.1092    0.0753    1.45    0.147
```

```

## Year2004          0.1078      0.0804      1.34      0.180
## Year2005          0.0559      0.0756      0.74      0.460
## Year2006          0.0905      0.0790      1.14      0.252
## Year2007          0.1214      0.0730      1.66      0.096 .
## Year2008          0.0739      0.0720      1.03      0.305
## Year2009          0.1328      0.0744      1.78      0.074 .
## Year2010          0.2001      0.0754      2.65      0.008 **
## Year2011          0.1400      0.0737      1.90      0.058 .
## Year2012          0.1071      0.0743      1.44      0.150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.006, Adjusted R-squared:  0.00138
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 320 weights are ~= 1. The remaining 3355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0118 0.8680 0.9490 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.027 1      1.013
## Year      1.027 16      1.001

```

## Residuals from last author



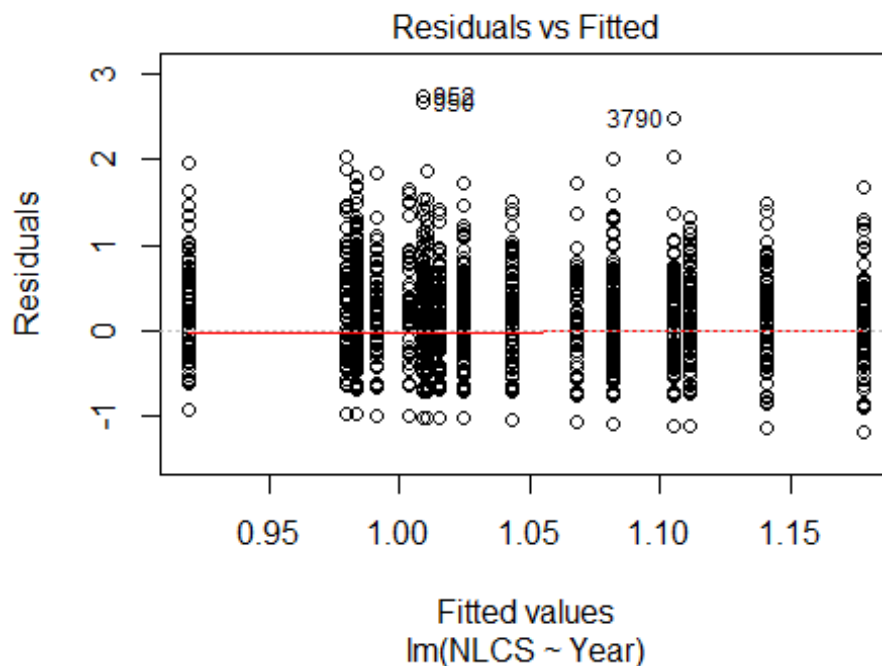
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1425  0001460136 3.678 2000    1703     4    2.611
## 1551  0006407254 3.743 2000    1703     4    2.676
## 1627  0001224048 3.749 2001    1702     4    2.710
## 8212  78650862532 3.769 2011    1804     2    2.660
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.16830 -0.40936  0.00491  0.41553  2.71058
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9689    0.0654   14.82  <2e-16 ***
## LastAuthorFemale1 -0.0179    0.0343   -0.52  0.6007
## Year1997        0.0603    0.0877    0.69  0.4918
## Year1998        0.1496    0.0965    1.55  0.1210
## Year1999        0.0864    0.0936    0.92  0.3559
## Year2000        0.0954    0.0964    0.99  0.3223
## Year2001        0.0695    0.0839    0.83  0.4076
## Year2002        0.1893    0.0765    2.47  0.0134 *
## Year2003        0.1082    0.0754    1.43  0.1516
```

```

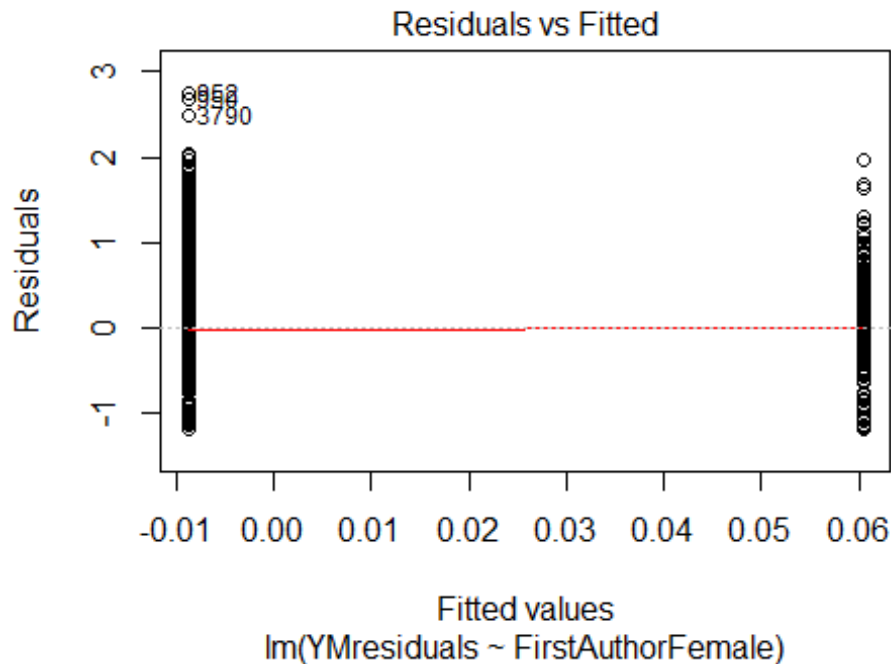
## Year2004          0.1069      0.0803      1.33      0.1833
## Year2005          0.0548      0.0757      0.72      0.4689
## Year2006          0.0886      0.0790      1.12      0.2621
## Year2007          0.1202      0.0731      1.65      0.0998 .
## Year2008          0.0724      0.0720      1.00      0.3150
## Year2009          0.1314      0.0744      1.77      0.0774 .
## Year2010          0.1994      0.0754      2.64      0.0082 **
## Year2011          0.1384      0.0738      1.88      0.0607 .
## Year2012          0.1064      0.0743      1.43      0.1526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.00598,    Adjusted R-squared:  0.00136
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 327 weights are ~= 1. The remaining 3348 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0117 0.8680 0.9490 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3675"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2614"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 152 162 173 167 236 227 197 136 116 145 220 248 256 252 182
## 2011 2012
## 185 178
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    66    74    86    90   128   116   109    88    63   105   133   182   173   166   125
## 2011 2012
##   123   134
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    64    71    81    87   117   112    97    80    55    98   122   157   150   152   109
## 2011 2012
##   115   117
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.07
```



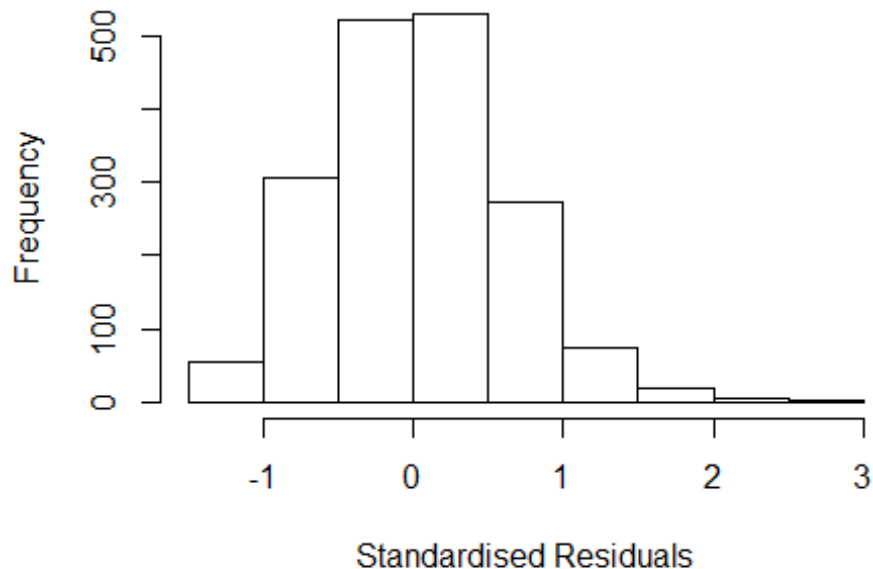
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.93376347822342"
## [1] "Male first author team size 2018 geometric mean: 1.99510736339862"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 980, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.07685534635435"
## [1] "Male last author team size 2018 geometric mean: 1.97270464637298"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 960, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.230 1          1.109
## LastAuthorFemale  1.229 1          1.109
## UniqueAuthors    1.210 4          1.024
## Year              1.306 16         1.008
```



## Residuals from first and last author and team size



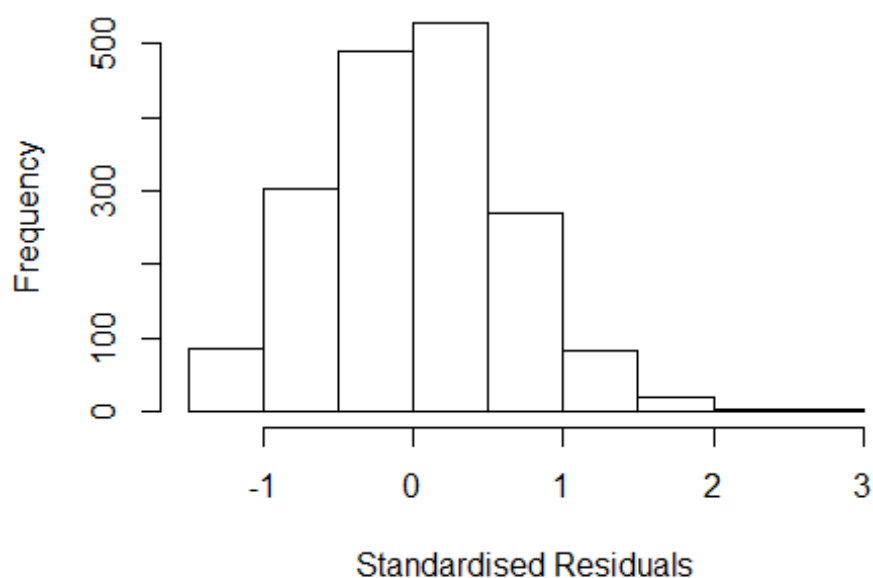
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 950   0001460136 3.678 2000    1703     4     2.548
## 952   0006407254 3.743 2000    1703     4     2.623
## 3790 79960204163 3.584 2011    1700     2     2.505
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.30636 -0.39922  0.00518  0.40309  2.62272
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.882817   0.097118   9.09  < 2e-16 ***
## FirstAuthorFemale1  0.062874   0.046486   1.35    0.18
## LastAuthorFemale1 -0.032972   0.047688  -0.69    0.49
## UniqueAuthors2     0.170412   0.034051   5.00 6.2e-07 ***
## UniqueAuthors3     0.334624   0.043671   7.66 3.0e-14 ***
## UniqueAuthors4     0.324853   0.077514   4.19 2.9e-05 ***
## UniqueAuthors5     0.494624   0.077668   6.37 2.4e-10 ***
## Year1997         -0.000865   0.130298  -0.01    0.99
```

```

## Year1998      -0.041337    0.128008    -0.32      0.75
## Year1999      0.020815    0.120423     0.17      0.86
## Year2000     -0.087387    0.113878    -0.77      0.44
## Year2001     -0.044135    0.119419    -0.37      0.71
## Year2002     -0.102762    0.122946    -0.84      0.40
## Year2003      0.141950    0.122538     1.16      0.25
## Year2004      0.106293    0.136019     0.78      0.43
## Year2005      0.126210    0.113001     1.12      0.26
## Year2006     -0.027788    0.108599    -0.26      0.80
## Year2007     -0.068168    0.109078    -0.62      0.53
## Year2008      0.052298    0.108587     0.48      0.63
## Year2009     -0.011278    0.107686    -0.10      0.92
## Year2010     -0.036713    0.109878    -0.33      0.74
## Year2011      0.026047    0.110483     0.24      0.81
## Year2012      0.050799    0.113229     0.45      0.65
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.0625, Adjusted R-squared:  0.0508
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1627 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0214 0.8530 0.9490 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.215 1 1.102
## LastAuthorFemale 1.208 1 1.099
## Year 1.080 16 1.002

```

## Residuals from first and last author



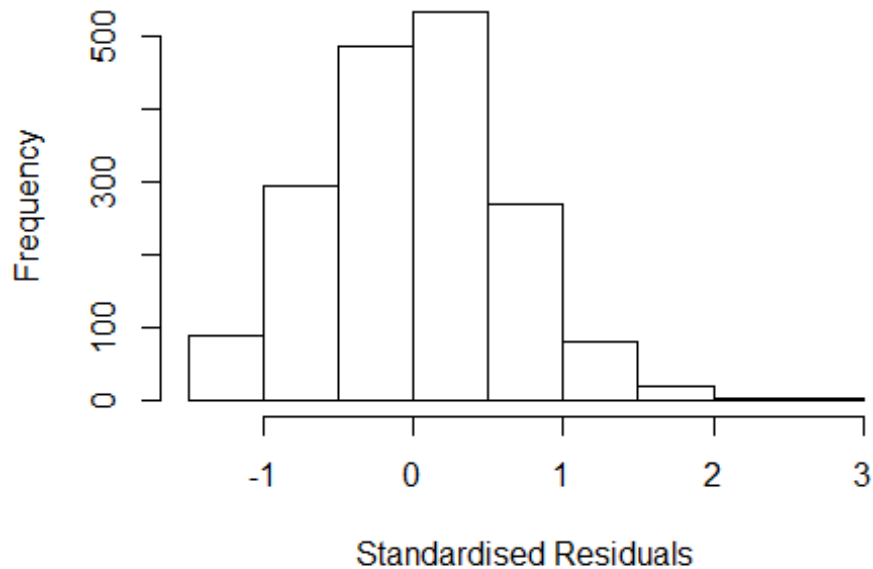
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 950   0001460136 3.678 2000    1703     4     2.756
## 952   0006407254 3.743 2000    1703     4     2.821
## 3790 79960204163 3.584 2011    1700     2     2.547
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16224 -0.41352  0.00733  0.42238  2.82109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.946038   0.095979   9.86  <2e-16 ***
## FirstAuthorFemale1 0.089250   0.047787   1.87   0.062 .
## LastAuthorFemale1 -0.019099   0.049972  -0.38   0.702
## Year1997         0.008219   0.129215   0.06   0.949
## Year1998        -0.020902   0.128775  -0.16   0.871
## Year1999         0.075993   0.122757   0.62   0.536
## Year2000        -0.024132   0.112914  -0.21   0.831
## Year2001         0.000187   0.118868   0.00   0.999
## Year2002        -0.028509   0.120430  -0.24   0.813
## Year2003         0.216204   0.120244   1.80   0.072 .
```

```

## Year2004          0.119591    0.137107    0.87    0.383
## Year2005          0.189373    0.114553    1.65    0.098 .
## Year2006          0.017636    0.108511    0.16    0.871
## Year2007         -0.000762    0.108618   -0.01    0.994
## Year2008          0.126452    0.107872    1.17    0.241
## Year2009          0.055298    0.106398    0.52    0.603
## Year2010          0.031516    0.109901    0.29    0.774
## Year2011          0.090750    0.109862    0.83    0.409
## Year2012          0.156578    0.112156    1.40    0.163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.621
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.00502
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1627 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0034 0.8640 0.9510 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1      1.019
## Year              1.038 16      1.001

```

## Residuals from first author



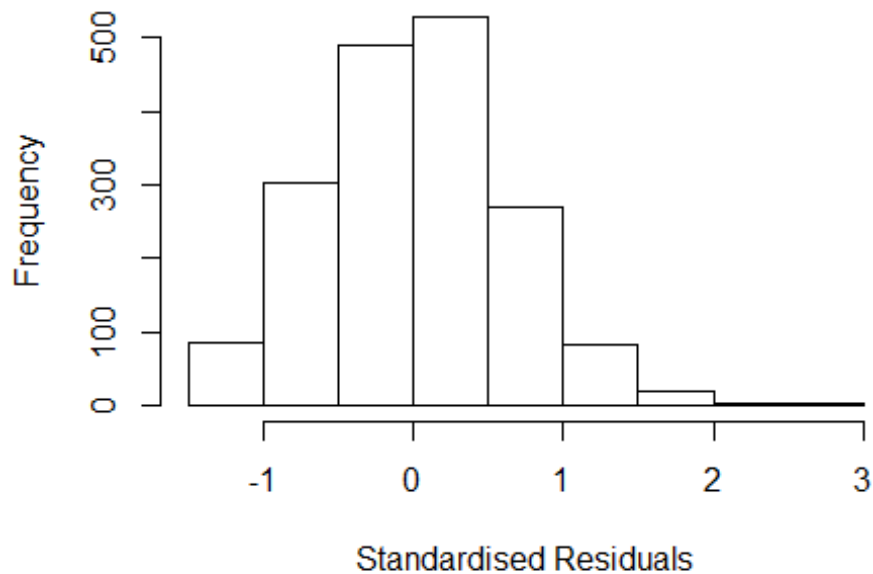
```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 950   0001460136 3.678 2000    1703     4     2.756
## 952   0006407254 3.743 2000    1703     4     2.821
## 3790 79960204163 3.584 2011    1700     2     2.547
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16098 -0.41197  0.00763  0.42167  2.82224
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.45e-01   9.58e-02   9.87  <2e-16 ***
## FirstAuthorFemale1  8.19e-02   4.43e-02   1.85   0.065 .
## Year1997         6.62e-03   1.29e-01   0.05   0.959
## Year1998        -2.12e-02   1.29e-01  -0.16   0.869
## Year1999         7.56e-02   1.23e-01   0.62   0.538
## Year2000        -2.43e-02   1.13e-01  -0.22   0.829
## Year2001        -3.47e-05   1.19e-01   0.00   1.000
## Year2002        -2.83e-02   1.20e-01  -0.24   0.814
## Year2003         2.16e-01   1.20e-01   1.80   0.072 .
## Year2004         1.18e-01   1.37e-01   0.86   0.388
```

```

## Year2005          1.88e-01  1.14e-01  1.64  0.100
## Year2006          1.83e-02  1.08e-01  0.17  0.866
## Year2007         -2.37e-03  1.08e-01 -0.02  0.983
## Year2008          1.26e-01  1.08e-01  1.17  0.242
## Year2009          5.55e-02  1.06e-01  0.52  0.602
## Year2010          3.14e-02  1.10e-01  0.29  0.775
## Year2011          9.08e-02  1.10e-01  0.83  0.408
## Year2012          1.57e-01  1.12e-01  1.40  0.162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.62
## Multiple R-squared:  0.015, Adjusted R-squared:  0.00553
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 156 weights are ~= 1. The remaining 1628 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0032 0.8650 0.9500 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1      1.017
## Year      1.035 16      1.001

```

## Residuals from last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 950   0001460136 3.678 2000    1703     4     2.756
## 952   0006407254 3.743 2000    1703     4     2.821
## 3790 79960204163 3.584 2011    1700     2     2.547
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.16244 -0.40966  0.00461  0.41510  2.81086
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.947945   0.095298   9.95  <2e-16 ***
## LastAuthorFemale1 0.018634   0.045702   0.41   0.684
## Year1997        0.015671   0.129089   0.12   0.903
## Year1998       -0.016489   0.128352  -0.13   0.898
## Year1999        0.079509   0.122218   0.65   0.515
## Year2000       -0.015805   0.112455  -0.14   0.888
## Year2001        0.011694   0.118368   0.10   0.921
## Year2002       -0.022750   0.119864  -0.19   0.849
## Year2003        0.214499   0.119533   1.79   0.073 .
## Year2004        0.118123   0.136851   0.86   0.388
```

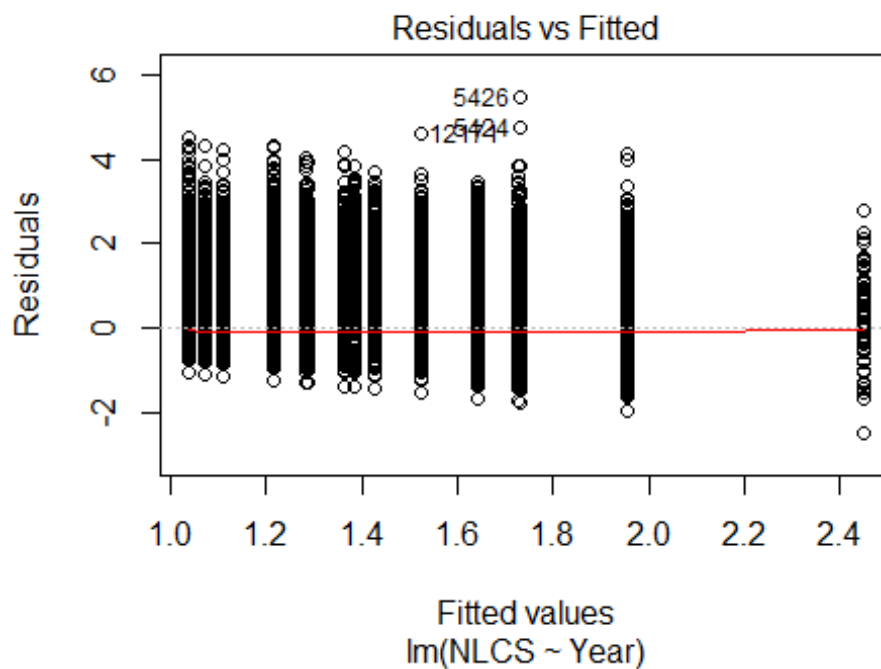
```

## Year2005      0.192886   0.113351   1.70   0.089 .
## Year2006      0.027711   0.107993   0.26   0.798
## Year2007     -0.000125   0.107974   0.00   0.999
## Year2008      0.128698   0.107125   1.20   0.230
## Year2009      0.063444   0.105544   0.60   0.548
## Year2010      0.036257   0.109496   0.33   0.741
## Year2011      0.102724   0.109406   0.94   0.348
## Year2012      0.162164   0.111399   1.46   0.146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.621
## Multiple R-squared:  0.0132, Adjusted R-squared:  0.00367
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1624 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0046 0.8650 0.9510 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1784"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2700"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2197 2298 2117 2018  220 2148 1880 1690 1720 1757 2048 2510 2825 3282 3231
## 2011 2012
## 2769 3211
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

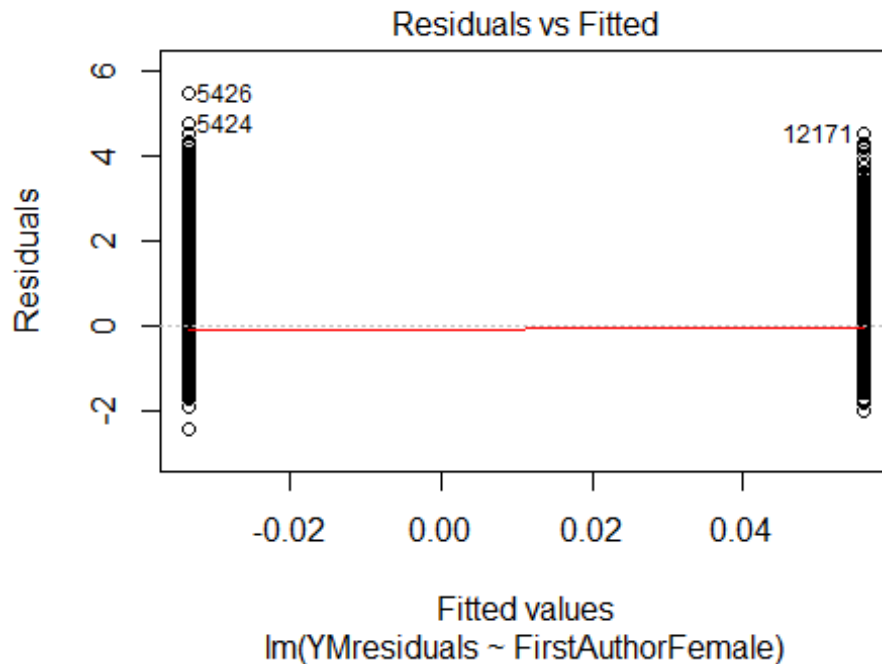
```



```
## 760 832 669 744 73 605 1139 1072 1119 1217 1375 1767 1999 2318 2181
## 2011 2012
## 1921 2330
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 705 764 619 665 65 564 1014 931 976 1080 1210 1581 1773 2024 1945
## 2011 2012
## 1717 2081
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 1000, df = 16, p-value <2e-16
```

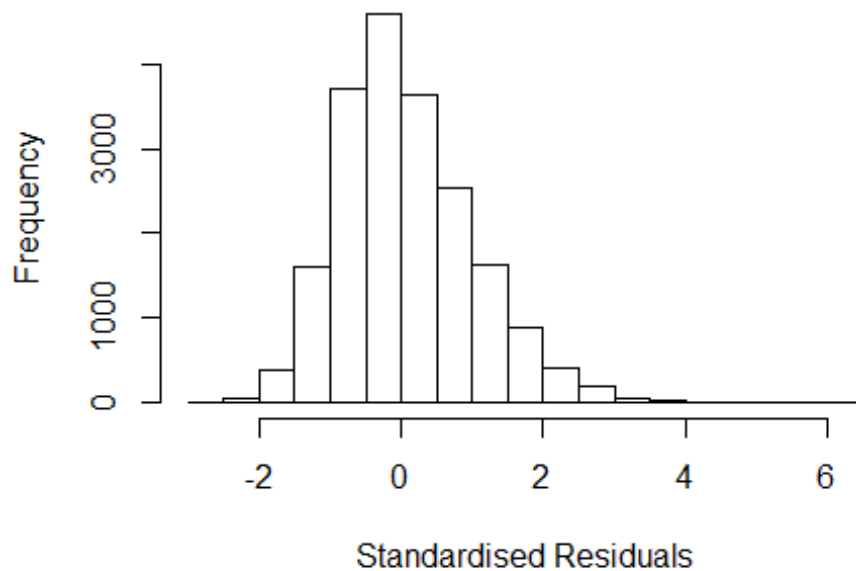


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.3, df = 1, p-value = 0.007
```



```
## [1] "Female first author team size 2018 geometric mean: 3.13830324052006"
## [1] "Male first author team size 2018 geometric mean: 2.93352973546078"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 220000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.83446368941595"
## [1] "Male last author team size 2018 geometric mean: 3.16471239407691"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180000, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.559 1      1.248
## LastAuthorFemale  1.573 1      1.254
## UniqueAuthors    1.094 4      1.011
## Year              1.085 16     1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 288 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 34      0029822959 3.838 1996      2700      1      2.531
## 55      0001945410 4.039 1996      2700      1      2.732
## 439     1542423634 0.000 1996      2700      2     -2.888
## 449     33746427540 0.000 1996      1300      2     -2.888
## 589     0030607603 6.092 1996      2700      1      3.241
## 596     0029808858 4.053 1996      2700      1      2.746
## 599     0029851867 3.948 1996      2700      1      2.641
## 613     0030599530 4.962 1996      2700      1      2.839
## 632     0030590742 4.215 1996      2700      1      2.908
## 729     0029805072 5.315 1996      2700      1      2.999
## 733     0029803181 4.808 1996      2700      1      3.501
## 1017    0029937218 5.035 1996      2700      1      2.757
## 1050    0030572468 4.501 1996      2700      1      3.194
## 1172    0029892771 4.222 1996      2700      1      2.915
## 1290    0029938999 4.371 1996      2700      1      2.994
## 1311    0029871557 4.455 1996      2700      1      3.078
## 1467    0029978223 4.025 1996      2700      1      2.648
## 1476    0029925181 5.956 1996      2700      1      3.121
## 1592    0030020685 4.440 1996      2700      1      2.544
## 1662    0029671048 4.411 1996      2700      1      2.515
## 1806    0029813554 4.211 1996      2700      1      2.834
## 2510    0030982113 3.783 1997      2700      1      2.795
## 2766    0031590608 4.218 1997      2700      1      2.625
## 2779    0030682092 4.703 1997      2700      1      2.690
## 2785    0030711655 4.498 1997      2700      1      2.922
```

## 2832	0031573250	5.072	1997	2700	1	2.557
## 2938	0030843904	4.847	1997	2700	1	2.989
## 3009	0030931611	4.657	1997	2700	1	2.714
## 3014	0030779141	4.141	1997	2700	1	2.565
## 3036	0030931456	4.323	1997	2700	1	3.265
## 3129	0030849415	4.518	1997	2700	1	2.942
## 3169	0342327375	4.622	1997	2700	1	3.634
## 3195	0030850011	3.794	1997	2700	1	2.736
## 3245	0030981444	4.075	1997	2700	1	3.087
## 3362	0031006994	4.369	1997	2700	1	2.739
## 3370	0030955988	4.114	1997	2700	1	3.126
## 3484	0342940785	4.741	1997	2700	1	3.753
## 3516	0030986151	4.459	1997	2700	1	2.516
## 3581	0008544037	5.001	1997	2700	1	4.013
## 3586	0030936556	4.929	1997	2700	1	3.071
## 3617	0030897243	3.553	1997	2700	1	2.565
## 3628	0030966012	4.618	1997	2700	1	2.675
## 3701	0030952324	4.196	1997	2700	1	2.566
## 3714	0031047332	4.445	1997	2700	1	2.815
## 3883	0031030622	3.772	1997	2700	1	2.714
## 3897	0031025486	4.377	1997	2700	1	2.801
## 4256	0030926893	3.492	1997	2700	1	2.504
## 5017	0032151752	4.597	1998	2700	1	3.560
## 5122	0032090387	0.000	1998	1304	3	-2.564
## 5293	0032556176	4.149	1998	2700	1	2.507
## 5297	0032541839	3.680	1998	2700	1	2.573
## 5369	0032189103	5.551	1998	2700	1	2.970
## 5424	0032511566	6.456	1998	2700	1	5.419
## 5426	0032511583	7.183	1998	2700	1	6.146
## 5431	7344263454	4.333	1998	2700	1	3.296
## 5442	0031822907	3.759	1998	2700	1	2.652
## 5594	0032497075	3.844	1998	2700	1	2.807
## 5711	0032474679	4.210	1998	2700	1	2.531
## 5770	0032550854	4.228	1998	2700	1	2.586
## 5876	0032580764	4.527	1998	2700	1	2.535
## 6169	0032492429	4.972	1998	2700	1	2.980
## 6232	0032574158	4.331	1998	2700	1	2.689
## 6329	0032583853	4.057	1998	2700	1	3.020
## 6389	0032501528	4.831	1998	2700	1	3.206
## 6599	0031798225	0.000	1998	2700	1	-2.564
## 7636	0033576162	4.142	1999	2700	1	2.534
## 7663	0033539389	4.128	1999	2700	1	3.038
## 7744	0033546995	5.299	1999	2700	1	3.621
## 8054	0000093848	5.112	1999	2700	1	2.549
## 8300	85007738351	0.000	1999	2700	1	-2.617
## 8466	0033608466	5.159	1999	2700	1	3.168
## 8584	0033616462	4.524	1999	2700	1	2.635
## 8600	0033596541	4.279	1999	2700	1	3.189
## 8855	0032904303	0.000	1999	2700	1	-2.547
## 9728	0343953074	4.490	2000	2700	1	2.733

##	9747	0034685157	4.545	2000	2700	1	2.788
##	10399	0035883865	4.405	2001	2700	1	2.736
##	10401	0035883898	4.507	2001	2700	1	2.977
##	10419	0035828388	4.911	2001	2700	1	2.670
##	10567	0035845283	4.839	2001	2700	1	2.544
##	10692	0034953155	3.289	2001	2700	1	2.575
##	10796	0035938977	4.761	2001	2700	1	3.231
##	10869	0035795969	5.211	2001	2700	1	4.497
##	11104	0035906301	4.056	2001	2700	1	2.526
##	11352	0035835406	3.975	2001	2700	1	3.261
##	11356	0035835455	4.430	2001	2700	1	2.900
##	11435	0035915621	4.792	2001	2700	1	4.078
##	11464	0035814611	4.835	2001	2700	1	2.540
##	12313	0037202790	5.023	2002	2700	1	2.685
##	12331	0037164375	5.152	2002	2700	1	3.563
##	12639	0036711316	4.157	2002	2700	1	2.584
##	12866	0037142944	3.529	2002	2700	1	2.772
##	12884	0037072062	4.256	2002	2700	1	2.857
##	13022	0037097683	4.661	2002	2700	1	2.949
##	13034	0037042471	3.964	2002	2700	1	2.565
##	13040	0037042674	4.421	2002	2700	1	2.709
##	13066	0036606736	4.768	2002	2700	1	3.941
##	13072	0001752768	4.604	2002	1300	2	3.847
##	13087	0037198418	4.789	2002	2700	1	2.505
##	13092	0037172392	4.111	2002	2700	1	2.766
##	13234	0037070772	3.480	2002	2700	1	2.653
##	13237	0037070811	4.227	2002	2700	1	3.400
##	13251	0037029292	4.622	2002	2700	1	2.995
##	13252	0037029304	4.335	2002	2700	1	2.607
##	13327	0037160919	4.208	2002	2700	1	2.619
##	13373	0036180760	3.622	2002	2700	1	2.865
##	13527	0037132827	3.557	2002	2700	1	2.730
##	13542	0037065542	3.750	2002	2700	1	2.923
##	14358	0345118958	4.063	2003	2700	1	2.554
##	14520	0142088505	4.874	2003	2700	1	3.225
##	14536	0141751673	3.918	2003	2700	1	2.636
##	14660	0141454739	3.358	2003	2700	1	2.664
##	14701	0042878401	3.944	2003	2700	1	2.608
##	14707	0041859261	5.128	2003	2700	1	2.853
##	14804	0042658340	4.480	2003	2700	1	2.815
##	14808	0043234290	4.784	2003	2700	1	2.563
##	14938	0038304773	4.450	2003	2700	1	3.686
##	15259	0037464519	3.414	2003	2700	1	2.720
##	15368	0037471825	3.966	2003	2700	1	2.630
##	15369	0037471833	4.376	2003	2700	1	2.727
##	15415	0037426064	4.867	2003	2700	1	2.646
##	15470	0037351456	3.238	2003	2700	2	2.544
##	15643	0037464756	4.600	2003	2700	1	2.898
##	15959	0642365217	3.249	2003	2700	1	2.555
##	16202	10644273470	3.874	2004	2700	1	3.226

##	16299	13244281317	4.541	2004	1315	2	3.304
##	16513	4844229917	3.968	2004	2700	1	2.677
##	16704	4344675264	4.089	2004	1704	3	2.625
##	16737	4344683381	5.510	2004	2700	1	3.280
##	16889	3242677099	4.820	2004	2700	1	2.590
##	16895	3242723235	4.333	2004	2700	1	2.714
##	16922	3042723720	5.214	2004	2700	1	3.038
##	17001	3042700097	5.251	2004	2700	1	3.632
##	17049	2942672794	3.342	2004	2700	1	2.624
##	17135	2442700599	4.355	2004	2700	1	2.698
##	17515	1542301591	3.968	2004	2700	1	2.504
##	17654	0842286066	3.353	2004	2700	1	2.705
##	17655	0842348097	4.802	2004	2700	1	3.495
##	18130	29544438304	4.009	2005	2700	1	2.781
##	18168	33646826151	3.107	2005	2700	1	2.521
##	18352	28044471557	3.284	2005	2700	1	2.628
##	18516	26244444318	4.188	2005	2700	1	3.014
##	18646	24744436009	3.317	2005	2700	1	2.661
##	18647	24744456592	3.458	2005	2700	1	2.872
##	19168	19744380352	4.678	2005	2700	1	4.092
##	19185	19344368841	5.322	2005	2700	1	3.193
##	19190	18444382649	3.463	2005	2700	1	2.877
##	19194	18744369929	4.646	2005	2700	1	4.060
##	19429	15244361216	5.216	2005	2700	1	4.630
##	19531	14644394273	4.111	2005	2700	1	3.525
##	19533	14644412384	3.452	2005	2700	1	2.866
##	19914	26944466840	5.136	2005	2700	1	3.023
##	19915	26944468736	4.069	2005	2700	1	2.614
##	20208	33845219790	3.707	2006	2700	1	3.054
##	20328	34347238420	3.322	2006	2700	1	2.669
##	20568	33750626215	4.171	2006	2700	1	2.860
##	20571	33750925581	4.007	2006	2700	1	2.539
##	20874	33748367018	4.269	2006	2700	1	3.028
##	20888	33748416499	5.205	2006	2700	1	3.009
##	21509	33744988382	3.169	2006	2700	1	2.516
##	21732	33645800971	3.258	2006	2700	1	2.605
##	22033	32644490529	4.692	2006	2700	1	2.512
##	22159	298444444176	4.173	2006	2700	1	2.932
##	22625	37049013701	5.171	2007	2700	1	3.016
##	22898	36749094450	3.747	2007	2700	1	2.547
##	22940	36348944504	5.457	2007	2700	1	3.891
##	22953	35548943929	4.812	2007	2700	1	2.673
##	23173	34548509498	4.045	2007	2700	1	2.656
##	23323	34548861255	3.930	2007	2700	1	2.730
##	23342	34548430156	3.639	2007	2700	1	3.081
##	23512	34548014479	3.324	2007	2700	1	2.766
##	23541	34547837402	4.813	2007	2700	1	2.728
##	23563	34547693119	5.506	2007	2700	1	3.405
##	23572	34547764305	4.959	2007	2700	1	2.874
##	23597	34547644454	5.013	2007	2700	1	2.928

##	23703	34447531794	3.425	2007	2700	1	2.797
##	23723	34447255724	4.649	2007	2700	1	2.564
##	23727	34447630803	4.674	2007	2700	1	2.535
##	23747	34447572979	4.812	2007	2700	1	3.596
##	23838	38449106012	3.101	2007	2700	1	2.543
##	23959	34250375564	4.631	2007	2700	1	2.530
##	23965	34250003794	4.837	2007	2700	1	2.698
##	24061	36248953716	4.289	2007	2700	1	3.089
##	24145	34249341210	4.103	2007	2700	1	2.730
##	24158	34248157118	4.876	2007	2700	1	3.730
##	24482	34347261376	3.220	2007	2700	1	2.662
##	24526	33847677585	4.250	2007	2700	1	2.877
##	24529	33947611404	3.302	2007	2700	1	2.674
##	24548	33947194241	4.947	2007	2700	1	3.435
##	24549	33947228816	4.850	2007	2700	1	2.749
##	24550	33947241772	4.874	2007	2700	1	2.773
##	24797	33847152253	4.577	2007	2700	1	3.065
##	24840	33846627856	4.213	2007	2700	1	3.655
##	24967	33846049673	4.834	2007	2700	1	2.749
##	25035	33846666462	3.169	2007	2700	1	2.611
##	25373	57649198083	3.905	2008	2700	1	3.443
##	25877	55149122038	5.100	2008	2700	1	3.111
##	25878	55149125165	4.787	2008	2700	1	3.737
##	26243	52249115660	3.717	2008	2700	1	3.255
##	26701	47149112621	5.086	2008	2700	1	3.097
##	26879	47149088261	4.789	2008	2700	1	2.746
##	26884	45949085378	4.364	2008	2700	1	2.948
##	27143	44449103216	4.014	2008	2700	1	2.581
##	27159	44249110469	3.172	2008	2700	1	2.710
##	27460	42149161021	4.109	2008	2700	1	2.832
##	27465	41949118582	2.993	2008	2700	1	2.531
##	27664	40949139693	4.069	2008	2700	1	2.653
##	27729	39849091058	3.927	2008	2700	1	2.650
##	28146	38049047178	5.329	2008	2700	1	3.340
##	28261	42949173113	3.065	2008	2700	1	2.533
##	28398	57849129354	3.110	2008	2700	1	2.648
##	28717	71749092363	3.897	2009	2700	1	2.910
##	28747	71549136477	3.762	2009	2700	1	2.564
##	29234	70450081001	5.303	2009	2700	1	3.447
##	29257	71249104673	3.549	2009	2700	1	2.562
##	29288	70449650655	4.865	2009	2700	1	3.009
##	29468	70350310101	3.801	2009	2700	1	3.472
##	29583	69249220030	2.840	2009	2700	1	2.511
##	29589	70349611684	5.029	2009	2700	1	3.173
##	29715	70350018319	5.363	2009	2700	1	3.491
##	29813	68849108622	2.995	2009	2700	1	2.666
##	29957	70350731580	2.855	2009	2700	1	2.526
##	29971	68349148204	3.480	2009	2700	1	2.563
##	30069	69449097272	3.961	2009	2700	1	2.677
##	30191	67651180816	4.434	2009	2700	1	2.578

##	30199	67650607983	5.537	2009	2700	1	3.681
##	30259	68149112409	4.126	2009	2700	1	3.155
##	30265	67649649676	4.676	2009	2700	1	3.759
##	30433	67649111182	4.381	2009	2700	1	2.525
##	30468	67650076181	3.734	2009	2700	1	2.520
##	30498	66649096725	3.506	2009	2700	1	3.177
##	30526	66149186737	3.525	2009	2700	1	2.554
##	30683	66249141245	4.629	2009	2700	1	2.757
##	30701	65649084159	5.220	2009	2700	1	3.882
##	30752	65649104423	3.285	2009	2700	1	2.956
##	30818	65349158356	3.757	2009	2700	1	2.840
##	31095	61849161041	3.506	2009	2700	1	2.535
##	31282	62849102089	3.467	2009	1000	2	2.550
##	31334	63149132663	3.294	2009	2700	1	2.895
##	31347	61749089425	2.982	2009	2700	1	2.583
##	31416	61649091385	4.600	2009	2700	1	2.744
##	31676	59049105205	2.855	2009	2700	1	2.526
##	31796	58749091313	2.914	2009	2700	1	2.585
##	31857	58749086000	4.008	2009	2700	1	2.864
##	32912	78649596720	3.725	2010	2700	1	2.527
##	33948	77956793539	3.509	2010	2700	1	2.576
##	34547	77953711938	3.886	2010	2700	1	2.899
##	34990	77951699612	4.010	2010	2700	1	3.681
##	35008	77951879165	4.717	2010	2700	1	2.807
##	35053	77950496251	2.867	2010	2700	1	2.538
##	35068	77950822827	4.845	2010	2700	1	2.989
##	35264	77749282912	4.078	2010	2700	1	2.794
##	35303	77749306261	4.234	2010	2700	1	3.074
##	35370	77950653061	2.950	2010	2700	1	2.621
##	35587	76749157063	3.860	2010	2700	1	2.716
##	35778	74849098087	3.649	2010	2700	1	3.320
##	35943	77949497387	3.679	2010	2700	1	2.708
##	37144	80855139395	2.967	2011	2700	1	2.598
##	37501	80052644219	3.458	2011	2700	1	3.089
##	37598	80052703117	4.390	2011	2700	1	4.021
##	37766	80052913150	4.505	2011	2700	1	2.539
##	37774	80051643493	3.245	2011	2700	1	2.876
##	37928	79956316608	3.223	2011	2700	1	2.784
##	37980	79960556064	3.432	2011	2700	1	3.063
##	38021	79958057463	3.014	2011	2700	1	2.645
##	38412	79955944811	3.589	2011	2700	1	2.632
##	38487	79955963939	3.644	2011	2700	1	2.687
##	38672	79955544813	4.917	2011	2700	1	3.733
##	38716	79953849295	3.530	2011	2700	1	3.091
##	38768	79953293021	4.330	2011	2700	1	2.952
##	39046	79952559080	3.319	2011	2700	1	2.950
##	39194	79951644557	4.908	2011	2700	1	2.958
##	39278	79551683853	3.319	2011	2700	1	2.950
##	39295	79952345398	2.999	2011	2700	1	2.560
##	39371	79251566511	3.650	2011	1000	2	2.693



```

## 39394 78651427507 4.543 2011      2700      1      2.647
## 40227 84863192356 2.999 2011      2700      1      2.630
## 40721 84879373604 3.020 2012      2700      1      2.722
## 41535 84865677850 4.010 2012      2700      1      2.897
## 41680 84865412684 3.020 2012      2700      1      2.722
## 42052 84863558979 3.490 2012      2700      1      2.534
## 42699 84860116942 3.548 2012      2700      1      3.250
## 42756 84859726718 3.810 2012      2700      1      2.870
## 42920 84859188542 3.964 2012      2700      1      2.658
## 43248 84857420826 4.143 2012      2700      1      2.976
## 43332 84856669252 4.530 2012      2700      1      2.705
## 43556 84859005251 3.744 2012      2700      1      2.631
## 43816 84863426446 5.368 2012      2700      1      3.489
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8885 -0.5860 -0.0795  0.6482  6.1458
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3073     0.0534   24.47 < 2e-16 ***
## FirstAuthorFemale1  0.0539     0.0170    3.18 0.00147 **
## LastAuthorFemale1  0.0162     0.0176    0.92 0.35792
## UniqueAuthors2     0.5882     0.0184   32.01 < 2e-16 ***
## UniqueAuthors3     0.8155     0.0203   40.07 < 2e-16 ***
## UniqueAuthors4     0.9549     0.0246   38.89 < 2e-16 ***
## UniqueAuthors5     1.5272     0.0235   64.91 < 2e-16 ***
## Year1997          -0.3192     0.0712   -4.48 7.3e-06 ***
## Year1998          -0.2701     0.0740   -3.65 0.00026 ***
## Year1999          -0.2878     0.0700   -4.11 3.9e-05 ***
## Year2000           0.4493     0.1894    2.37 0.01768 *
## Year2001          -0.5932     0.0697   -8.51 < 2e-16 ***
## Year2002          -0.5502     0.0624   -8.82 < 2e-16 ***
## Year2003          -0.6137     0.0633   -9.70 < 2e-16 ***
## Year2004          -0.6589     0.0617  -10.68 < 2e-16 ***
## Year2005          -0.7214     0.0594  -12.15 < 2e-16 ***
## Year2006          -0.6546     0.0591  -11.08 < 2e-16 ***
## Year2007          -0.7498     0.0579  -12.95 < 2e-16 ***
## Year2008          -0.8458     0.0562  -15.06 < 2e-16 ***
## Year2009          -0.9786     0.0559  -17.49 < 2e-16 ***
## Year2010          -0.9786     0.0556  -17.61 < 2e-16 ***
## Year2011          -0.9385     0.0560  -16.75 < 2e-16 ***
## Year2012          -1.0097     0.0552  -18.31 < 2e-16 ***

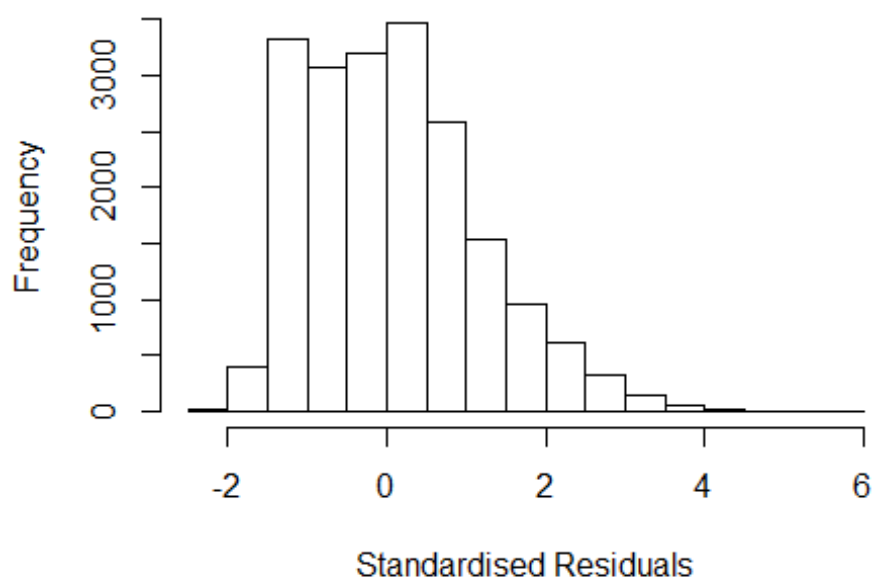
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.836
## Multiple R-squared:  0.304, Adjusted R-squared:  0.303
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 10 observations c(1187,1638,1639,3142,3285,3770,6855,6874,6979,16314)
## are outliers with |weight| = 0 ( < 5.1e-06);
## 1200 weights are ~= 1. The remaining 18504 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8590 0.9460 0.8900 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      5.07e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.404 1 1.185
## LastAuthorFemale 1.404 1 1.185
## Year 1.015 16 1.000

```

## Residuals from first and last author



```
## [1] "List of 542 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 90      0029925223 4.892 1996      2700      1      3.196
## 589     0030607603 6.092 1996      2700      1      4.396
## 613     0030599530 4.962 1996      2700      1      3.098
## 729     0029805072 5.315 1996      2700      1      3.275
## 733     0029803181 4.808 1996      2700      1      2.944
## 773     0029809469 4.580 1996      2700      1      2.716
## 852     0029785701 4.390 1996      2700      1      2.694
## 858     0029796404 4.540 1996      1300      2      2.676
## 1017    0029937218 5.035 1996      2700      1      3.339
## 1040    0030582181 4.488 1996      2700      1      2.624
## 1050    0030572468 4.501 1996      2700      1      2.637
## 1299    0030007093 4.561 1996      2700      1      2.697
## 1311    0029871557 4.455 1996      2700      1      2.583
## 1454    0029988543 4.496 1996      2700      1      2.632
## 1476    0029925181 5.956 1996      2700      1      4.092
## 1518    0030057755 4.693 1996      2700      1      2.829
## 1592    0030020685 4.440 1996      2700      1      2.576
## 1606    0030048806 4.899 1996      2700      1      2.859
## 1662    0029671048 4.411 1996      2700      1      2.547
## 1690    0030068535 4.699 1996      2700      1      3.003
## 2442    0030731968 4.336 1997      2700      1      2.857
## 2561    2642653233 4.177 1997      2700      1      2.698
## 2766    0031590608 4.218 1997      2700      1      2.907
## 2779    0030682092 4.703 1997      2700      1      3.216
## 2785    0030711655 4.498 1997      2700      1      3.019
```

##	2787	0031590552	4.556	1997	2700	1	3.077
##	2832	0031573250	5.072	1997	2700	1	3.593
##	2845	0030710196	4.233	1997	2700	1	2.922
##	2931	0030699243	4.669	1997	2700	1	3.190
##	2938	0030843904	4.847	1997	2700	1	3.192
##	2959	0030825804	4.114	1997	2700	1	2.635
##	2962	0030870938	4.251	1997	2700	1	2.940
##	2990	0030886421	3.902	1997	2700	1	2.591
##	2991	0030923557	4.251	1997	2700	1	2.596
##	3009	0030931611	4.657	1997	2700	1	3.178
##	3014	0030779141	4.141	1997	2700	1	2.662
##	3018	0030882421	4.784	1997	2700	1	3.129
##	3036	0030931456	4.323	1997	2700	1	2.836
##	3049	0030738896	4.082	1997	2700	1	2.771
##	3126	0030800984	4.242	1997	2700	1	2.763
##	3129	0030849415	4.518	1997	2700	1	3.039
##	3169	0342327375	4.622	1997	2700	1	3.143
##	3245	0030981444	4.075	1997	2700	1	2.596
##	3278	0030806335	3.985	1997	2700	1	2.506
##	3362	0031006994	4.369	1997	2700	1	2.714
##	3370	0030955988	4.114	1997	2700	1	2.635
##	3484	0342940785	4.741	1997	2700	1	3.262
##	3516	0030986151	4.459	1997	2700	1	2.980
##	3581	0008544037	5.001	1997	2700	1	3.522
##	3586	0030936556	4.929	1997	2700	1	3.274
##	3611	0030775107	4.041	1997	2700	1	2.562
##	3613	0030949484	4.274	1997	2700	1	2.795
##	3628	0030966012	4.618	1997	2700	1	3.139
##	3688	0030991825	4.401	1997	2700	1	2.922
##	3701	0030952324	4.196	1997	2700	1	2.541
##	3714	0031047332	4.445	1997	2700	1	2.790
##	3802	0031054894	4.060	1997	2700	1	2.581
##	3817	0031016791	4.320	1997	2700	1	2.841
##	3820	0031024016	3.830	1997	2700	1	2.519
##	3845	0031014243	4.141	1997	2700	1	2.662
##	3863	0031018731	4.631	1997	2700	1	3.152
##	3897	0031025486	4.377	1997	2700	1	2.898
##	4250	0030914797	4.320	1997	2700	1	2.665
##	4951	0032564695	4.550	1998	2700	1	2.975
##	5017	0032151752	4.597	1998	2700	1	3.022
##	5293	0032556176	4.149	1998	2700	1	2.742
##	5318	0032505605	4.331	1998	2700	1	2.924
##	5369	0032189103	5.551	1998	2700	1	4.144
##	5424	0032511566	6.456	1998	2700	1	4.881
##	5426	0032511583	7.183	1998	2700	1	5.608
##	5431	7344263454	4.333	1998	2700	1	2.758
##	5447	0032486979	4.238	1998	2700	1	2.663
##	5458	0032480253	4.870	1998	2700	1	3.463
##	5525	0032578283	4.135	1998	2700	1	2.560
##	5621	0032147105	4.300	1998	2700	1	2.893

## 5637	0032566160	4.969	1998	2700	1	3.394
## 5664	0032543934	4.213	1998	2700	1	2.638
## 5684	0032508294	4.995	1998	2700	1	3.588
## 5770	0032550854	4.228	1998	2700	1	2.821
## 5876	0032580764	4.527	1998	2700	1	2.952
## 5942	0032474225	4.624	1998	2700	1	3.049
## 5943	0032474229	4.290	1998	2700	1	2.715
## 6132	0032521142	4.442	1998	2700	1	2.867
## 6169	0032492429	4.972	1998	2700	1	3.397
## 6232	0032574158	4.331	1998	2700	1	2.924
## 6389	0032501528	4.831	1998	2700	1	3.256
## 6410	0032477529	4.501	1998	2700	1	2.750
## 7511	0033552391	4.335	1999	2700	1	2.752
## 7513	0033552424	4.437	1999	2700	1	2.678
## 7636	0033576162	4.142	1999	2700	1	2.559
## 7663	0033539389	4.128	1999	2700	1	2.537
## 7728	0033581376	4.083	1999	2700	1	2.500
## 7744	0033546995	5.299	1999	2700	1	3.708
## 7760	0033523401	4.771	1999	2700	1	3.180
## 7913	0033620867	4.158	1999	2700	1	2.567
## 7948	0033578120	4.184	1999	2700	1	2.601
## 8054	0000093848	5.112	1999	2700	1	3.697
## 8056	0033606486	4.150	1999	2700	1	2.567
## 8061	0033606539	4.310	1999	2700	1	2.727
## 8076	0033583970	4.325	1999	2700	1	2.566
## 8206	0033535768	4.297	1999	2700	1	2.882
## 8250	0033135998	4.757	1999	2700	1	3.342
## 8366	0033608862	4.261	1999	2700	1	2.678
## 8466	0033608466	5.159	1999	2700	1	3.744
## 8482	0033056940	4.126	1999	2700	1	2.543
## 8519	0033550718	4.179	1999	2700	1	2.596
## 8584	0033616462	4.524	1999	2700	1	2.765
## 8598	0033026308	4.715	1999	2700	1	3.132
## 8600	0033596541	4.279	1999	2700	1	2.688
## 8633	0033537299	4.467	1999	2700	1	2.884
## 8637	0033537344	4.827	1999	2700	1	3.068
## 8652	0033513984	4.145	1999	2700	1	2.730
## 8746	0032741972	4.414	1999	2700	1	2.655
## 9856	0035915319	3.646	2001	2700	1	2.509
## 9949	0035655197	4.230	2001	2700	1	3.093
## 9953	0035657149	4.504	2001	2700	1	3.359
## 9955	0035657752	4.736	2001	2700	1	3.591
## 10180	0035802719	4.271	2001	2700	1	3.134
## 10343	0035968604	3.781	2001	2700	1	2.636
## 10399	0035883865	4.405	2001	2700	1	3.268
## 10401	0035883898	4.507	2001	2700	1	3.370
## 10419	0035828388	4.911	2001	2700	1	3.774
## 10421	0035828406	4.115	2001	2700	1	2.802
## 10423	0035828439	3.858	2001	2700	1	2.545
## 10511	0035949107	4.065	2001	2700	1	2.928

##	10515	0035949122	3.764	2001	2700	1	2.627
##	10539	0035908713	3.917	2001	2700	1	2.780
##	10566	0035845259	4.459	2001	2700	1	3.490
##	10567	0035845283	4.839	2001	2700	1	3.526
##	10569	0035845310	4.139	2001	2700	1	3.002
##	10667	0035928402	4.299	2001	2700	1	3.162
##	10669	0035928414	4.026	2001	2700	1	3.057
##	10707	0035859468	3.873	2001	2700	1	2.560
##	10708	0035859488	3.903	2001	2700	1	2.766
##	10731	0035822265	4.099	2001	2700	1	2.962
##	10733	0035822274	4.398	2001	2700	1	3.261
##	10791	0035938907	3.794	2001	2700	1	2.657
##	10794	0035938954	4.242	2001	2700	1	3.105
##	10796	0035938977	4.761	2001	2700	1	3.624
##	10849	0035832508	3.881	2001	2700	1	2.568
##	10869	0035795969	5.211	2001	2700	1	4.074
##	10941	0035912582	4.041	2001	2700	1	3.072
##	10967	0035849287	4.303	2001	2700	1	3.166
##	10968	0035849309	3.630	2001	2700	1	2.661
##	10973	0035849348	3.798	2001	2700	1	2.661
##	11104	0035906301	4.056	2001	2700	1	2.919
##	11141	0035820302	3.681	2001	2700	1	2.536
##	11195	0035962071	4.328	2001	2700	1	3.191
##	11270	0035799105	4.367	2001	2700	1	3.054
##	11279	0035283198	3.917	2001	2700	1	2.780
##	11300	0035941743	4.099	2001	2700	1	2.962
##	11352	0035835406	3.975	2001	2700	1	2.838
##	11356	0035835455	4.430	2001	2700	1	3.293
##	11357	0035835456	4.299	2001	2700	1	3.162
##	11358	0035835465	3.760	2001	2700	1	2.623
##	11434	0035915614	4.165	2001	2700	1	3.020
##	11435	0035915621	4.792	2001	2700	1	3.655
##	11455	0035852473	3.982	2001	2700	1	2.669
##	11464	0035814611	4.835	2001	2700	1	3.522
##	12307	0037202686	4.381	2002	2700	1	3.039
##	12313	0037202790	5.023	2002	2700	1	3.506
##	12331	0037164375	5.152	2002	2700	1	3.978
##	12463	0037180219	3.877	2002	2700	1	2.535
##	12465	0037180235	4.153	2002	2700	1	2.811
##	12474	0037168030	4.107	2002	2700	1	2.590
##	12517	0037027198	4.571	2002	2700	1	3.221
##	12608	0037151919	4.032	2002	2700	1	2.690
##	12639	0036711316	4.157	2002	2700	1	2.815
##	12707	0037194703	3.922	2002	2700	1	2.572
##	12749	0037125595	4.180	2002	2700	1	2.663
##	12850	0037183247	4.135	2002	2700	1	2.793
##	12884	0037072062	4.256	2002	2700	1	2.739
##	13012	0037142007	4.465	2002	2700	1	3.123
##	13022	0037097683	4.661	2002	2700	1	3.319
##	13040	0037042674	4.421	2002	2700	1	3.079

##	13047	0037030659	4.332	2002	2700	1	2.815
##	13066	0036606736	4.768	2002	2700	1	3.418
##	13072	0001752768	4.604	2002	1300	2	3.262
##	13087	0037198418	4.789	2002	2700	1	3.447
##	13092	0037172392	4.111	2002	2700	1	2.769
##	13093	0037172426	4.590	2002	2700	1	3.248
##	13219	0037140185	4.238	2002	2700	1	2.721
##	13221	0037140211	4.025	2002	2700	1	2.508
##	13237	0037070811	4.227	2002	2700	1	2.877
##	13251	0037029292	4.622	2002	2700	1	3.105
##	13252	0037029304	4.335	2002	2700	1	3.161
##	13292	0037196927	3.943	2002	2700	1	2.593
##	13327	0037160919	4.208	2002	2700	1	3.034
##	13343	0037117094	4.013	2002	2700	1	2.671
##	13382	0037045846	4.216	2002	2700	1	2.874
##	13430	0037160756	4.507	2002	2700	1	3.157
##	13448	0037116793	3.905	2002	2700	1	2.731
##	13452	0037116856	4.079	2002	2700	1	2.905
##	13528	0037132846	4.730	2002	2700	1	3.388
##	14358	0345118958	4.063	2003	2700	1	2.779
##	14362	0345604401	4.243	2003	2700	1	2.959
##	14398	0242654882	4.568	2003	2700	1	3.284
##	14492	0242285681	4.228	2003	2700	1	2.944
##	14511	0142186278	4.266	2003	2700	1	2.982
##	14520	0142088505	4.874	2003	2700	1	3.590
##	14536	0141751673	3.918	2003	2700	1	2.634
##	14640	0141757403	3.972	2003	2700	1	2.512
##	14707	0041859261	5.128	2003	2700	1	3.668
##	14726	0141612000	4.003	2003	2700	1	2.719
##	14737	0142186681	4.254	2003	2700	1	2.794
##	14789	0042536436	4.076	2003	2700	1	2.616
##	14804	0042658340	4.480	2003	2700	1	3.363
##	14808	0043234290	4.784	2003	2700	1	3.500
##	14923	0038340259	4.057	2003	2700	1	2.597
##	14933	0037629268	3.791	2003	2700	1	2.507
##	14938	0038304773	4.450	2003	2700	1	3.157
##	14951	0038501052	4.215	2003	2700	1	2.931
##	15269	0037454376	3.950	2003	2700	1	2.833
##	15278	0037433669	3.825	2003	2700	1	2.532
##	15292	0037420477	3.972	2003	2700	1	2.855
##	15368	0037471825	3.966	2003	2700	1	2.506
##	15369	0037471833	4.376	2003	2700	1	3.092
##	15381	0037460811	4.023	2003	2700	1	2.739
##	15399	0037443878	4.020	2003	2700	1	2.736
##	15403	0037444242	3.641	2003	2700	1	2.524
##	15415	0037426064	4.867	2003	2700	1	3.583
##	15522	0037442694	4.176	2003	2700	1	2.883
##	15536	0037425770	4.632	2003	2700	1	3.348
##	15566	0037299629	3.641	2003	2700	1	2.524
##	15641	0037464751	3.818	2003	2700	1	2.534

##	15643	0037464756	4.600	2003	2700	1	3.140
##	15661	0037452226	3.862	2003	2700	1	2.578
##	15668	0037431695	4.060	2003	2700	1	2.767
##	15671	0037431719	3.844	2003	2700	1	2.727
##	15945	0347600946	4.318	2003	2700	1	3.034
##	16179	10944259169	3.825	2004	2700	1	2.591
##	16190	10344265934	4.562	2004	2700	1	3.328
##	16202	10644273470	3.874	2004	2700	1	2.640
##	16299	13244281317	4.541	2004	1315	2	3.307
##	16383	8544279621	4.120	2004	2700	1	2.886
##	16461	8444248351	3.914	2004	2700	1	2.680
##	16513	4844229917	3.968	2004	2700	1	2.558
##	16590	4644269881	3.881	2004	2700	1	2.647
##	16704	4344675264	4.089	2004	1704	3	2.855
##	16737	4344683381	5.510	2004	2700	1	4.100
##	16765	4043128084	3.950	2004	2700	1	2.716
##	16767	4043157677	4.539	2004	2700	1	3.305
##	16865	3242768475	4.240	2004	2700	1	2.997
##	16868	3542990935	3.776	2004	2700	1	2.542
##	16889	3242677099	4.820	2004	2700	1	3.410
##	16895	3242723235	4.333	2004	2700	1	3.266
##	16896	3242727457	3.954	2004	2700	1	2.887
##	16922	3042723720	5.214	2004	2700	1	3.980
##	17001	3042700097	5.251	2004	2700	1	4.184
##	17019	2942670131	4.495	2004	2700	1	3.261
##	17135	2442700599	4.355	2004	2700	1	2.945
##	17144	2442454953	3.773	2004	2700	1	2.539
##	17157	2342592457	3.836	2004	2700	1	2.769
##	17447	1542283597	3.988	2004	2700	1	2.745
##	17515	1542301591	3.968	2004	2700	1	2.734
##	17516	1542331537	3.852	2004	2700	1	2.618
##	17529	1342308180	4.448	2004	2700	1	3.038
##	17655	0842348097	4.802	2004	2700	1	3.559
##	18126	29544444716	4.109	2005	2700	1	2.970
##	18130	29544438304	4.009	2005	2700	1	2.695
##	18166	28844504752	4.453	2005	2700	1	3.314
##	18184	28244442441	4.424	2005	2700	1	3.285
##	18186	282444481060	3.857	2005	2700	1	2.718
##	18391	27744471516	3.860	2005	2700	1	2.546
##	18516	26244444318	4.188	2005	2700	1	3.049
##	18602	25144502474	3.767	2005	2700	1	2.628
##	18678	24044487679	3.984	2005	2700	1	2.670
##	18799	23444461703	4.567	2005	2700	1	3.420
##	19062	20444410110	3.807	2005	2700	1	2.668
##	19065	20444477045	3.829	2005	2700	1	2.690
##	19168	19744380352	4.678	2005	2700	1	3.539
##	19185	19344368841	5.322	2005	2700	1	4.351
##	19194	18744369929	4.646	2005	2700	1	3.507
##	19265	18044388395	4.220	2005	2700	1	2.906
##	19314	17244373777	3.838	2005	1000	2	2.524



##	19413	20144387123	4.557	2005	2700	1	3.410
##	19429	15244361216	5.216	2005	2700	1	4.077
##	19457	20044365983	4.316	2005	2700	1	3.177
##	19531	14644394273	4.111	2005	2700	1	2.972
##	19550	19744382704	3.823	2005	2700	1	2.684
##	19674	12344278985	3.664	2005	2700	1	2.525
##	19914	26944466840	5.136	2005	2700	1	3.997
##	19915	26944468736	4.069	2005	2700	1	2.755
##	20171	33845655265	3.983	2006	2700	1	2.787
##	20190	33845928026	4.180	2006	2700	1	2.984
##	20207	33845216421	4.276	2006	2700	1	3.072
##	20208	33845219790	3.707	2006	2700	1	2.511
##	20517	33845269095	3.712	2006	2700	1	2.516
##	20531	33751337383	4.059	2006	2700	1	2.863
##	20568	33750626215	4.171	2006	2700	1	2.967
##	20571	33750925581	4.007	2006	2700	1	2.811
##	20577	33750500333	4.675	2006	2700	1	3.303
##	20713	33749626549	3.883	2006	2700	1	2.687
##	20717	33750616076	3.876	2006	2700	1	2.504
##	20762	33750876203	3.908	2006	2700	1	2.712
##	20857	33748476238	3.824	2006	2700	1	2.628
##	20874	33748367018	4.269	2006	2700	1	3.073
##	20888	33748416499	5.205	2006	2700	1	4.176
##	20917	33748296240	4.269	2006	2700	1	2.897
##	21004	33747758047	4.019	2006	2700	1	2.647
##	21158	33746384604	4.618	2006	2700	1	3.422
##	21171	33746304156	4.304	2006	2700	1	3.100
##	21193	33745825433	4.090	2006	2700	1	2.894
##	21207	33745615112	3.747	2006	2700	1	2.551
##	21309	33745658837	4.213	2006	2700	1	2.841
##	21317	33745440706	3.989	2006	2700	1	2.793
##	21320	33745727486	3.584	2006	2700	1	2.555
##	21505	33646786398	4.166	2006	2700	1	2.970
##	21541	33646674466	4.119	2006	2700	1	2.923
##	21773	33645728134	4.034	2006	2700	1	2.838
##	21891	33644837550	3.725	2006	2700	1	2.529
##	22001	33344466383	4.144	2006	2700	1	2.948
##	22033	32644490529	4.692	2006	2700	1	3.496
##	22051	31844439769	4.084	2006	2700	1	2.888
##	22123	31444432059	3.854	2006	2700	1	2.658
##	22159	29844444176	4.173	2006	2700	1	2.977
##	22625	37049013701	5.171	2007	2700	1	4.088
##	22627	37049030806	3.471	2007	2700	1	2.564
##	22646	38049023073	4.141	2007	2700	1	3.067
##	22909	36148996837	3.610	2007	2700	1	2.703
##	22940	36348944504	5.457	2007	2700	1	4.207
##	22953	35548943929	4.812	2007	2700	1	3.562
##	22986	36849054594	3.515	2007	2700	1	2.608
##	22993	38149137150	3.935	2007	2700	1	3.028
##	23113	35548987593	3.973	2007	2700	1	2.890

##	23115	35349011537	4.544	2007	2700	1	3.470
##	23141	35248901000	4.436	2007	2700	1	3.362
##	23165	34948834261	4.114	2007	2700	1	3.207
##	23173	34548509498	4.045	2007	2700	1	3.138
##	23272	35148842531	4.087	2007	2700	1	3.180
##	23322	34548859010	3.950	2007	2700	1	2.867
##	23323	34548861255	3.930	2007	2700	1	2.680
##	23342	34548430156	3.639	2007	2700	1	2.565
##	23343	34548616859	3.747	2007	2700	1	2.664
##	23407	34548513296	3.771	2007	2700	1	2.688
##	23541	34547837402	4.813	2007	2700	1	3.739
##	23546	34548175719	4.103	2007	2700	1	3.029
##	23563	34547693119	5.506	2007	2700	1	4.599
##	23572	34547764305	4.959	2007	2700	1	3.885
##	23597	34547644454	5.013	2007	2700	1	3.939
##	23692	34547148986	3.610	2007	2700	1	2.536
##	23723	34447255724	4.649	2007	2700	1	3.575
##	23727	34447630803	4.674	2007	2700	1	3.424
##	23736	34447319145	4.410	2007	2700	1	3.336
##	23747	34447572979	4.812	2007	2700	1	3.729
##	23766	34447126847	4.160	2007	2700	1	3.086
##	23767	34447128057	3.688	2007	2700	1	2.614
##	23881	34447105600	3.765	2007	2700	1	2.515
##	23959	34250375564	4.631	2007	2700	1	3.724
##	23965	34250003794	4.837	2007	2700	1	3.587
##	24006	34347235863	3.646	2007	2700	1	2.572
##	24061	36248953716	4.289	2007	2700	1	3.039
##	24131	34249822239	3.715	2007	2700	1	2.632
##	24143	34249281687	3.954	2007	2700	1	2.880
##	24145	34249341210	4.103	2007	2700	1	3.029
##	24158	34248157118	4.876	2007	2700	1	3.802
##	24388	34047244736	3.702	2007	2700	1	2.795
##	24411	34247190031	3.702	2007	2700	1	2.628
##	24526	33847677585	4.250	2007	2700	1	3.176
##	24530	34247185441	4.298	2007	2700	1	3.215
##	24548	33947194241	4.947	2007	2700	1	3.873
##	24549	33947228816	4.850	2007	2700	1	3.943
##	24550	33947241772	4.874	2007	2700	1	3.967
##	24603	33947507103	4.263	2007	2700	1	3.180
##	24797	33847152253	4.577	2007	2700	1	3.503
##	24840	33846627856	4.213	2007	2700	1	3.139
##	24943	33846461161	4.331	2007	2700	1	3.257
##	24967	33846049673	4.834	2007	2700	1	3.760
##	24995	33846193221	3.973	2007	2700	1	2.899
##	25373	57649198083	3.905	2008	2700	1	2.898
##	25424	57749121512	4.445	2008	2700	1	3.262
##	25876	55149121818	3.553	2008	2700	1	2.713
##	25877	55149122038	5.100	2008	2700	1	4.093
##	25878	55149125165	4.787	2008	2700	1	3.780
##	25900	54949131747	4.309	2008	2700	1	3.126

##	25920	54349122201	3.593	2008	2700	1	2.586
##	25999	54849408740	3.510	2008	2700	1	2.503
##	26110	52649159199	3.642	2008	2700	1	2.626
##	26127	55849125109	4.199	2008	2700	1	3.016
##	26243	52249115660	3.717	2008	2700	1	2.710
##	26273	51049101957	3.610	2008	2700	1	2.603
##	26465	50449106919	3.416	2008	2700	1	2.576
##	26522	50149104801	3.785	2008	2700	1	2.778
##	26532	49249119141	4.376	2008	2700	1	3.369
##	26595	50949084462	3.740	2008	2700	1	2.557
##	26598	50949134384	3.348	2008	2700	1	2.508
##	26632	48249085691	3.364	2008	2700	3	2.524
##	26643	48249113707	4.179	2008	2700	1	2.996
##	26701	47149112621	5.086	2008	2700	1	4.079
##	26705	47549085541	3.510	2008	2700	1	2.503
##	26735	46949105321	4.336	2008	2700	1	3.329
##	26753	46049113677	3.962	2008	2700	1	2.779
##	26879	47149088261	4.789	2008	2700	1	3.606
##	26884	45949085378	4.364	2008	2700	1	3.357
##	27113	44849100254	3.626	2008	2700	1	2.786
##	27143	44449103216	4.014	2008	2700	1	3.174
##	27144	44449126657	3.806	2008	2700	1	2.623
##	27173	43449092540	3.839	2008	2700	1	2.832
##	27202	43749094124	3.688	2008	2700	1	2.672
##	27417	42549120016	3.688	2008	2700	1	2.848
##	27443	42249108710	3.776	2008	2700	1	2.593
##	27460	42149161021	4.109	2008	2700	1	3.102
##	27512	41849107583	4.348	2008	2700	1	3.165
##	27525	43249092442	3.668	2008	2700	1	2.661
##	27664	40949139693	4.069	2008	2700	1	3.062
##	27729	39849091058	3.927	2008	2700	1	2.920
##	27751	41549104622	3.722	2008	2700	1	2.706
##	27755	41549160713	4.051	2008	2700	1	3.044
##	27867	39649105037	4.378	2008	2700	1	3.371
##	28081	38149029249	3.437	2008	2700	1	2.597
##	28146	38049047178	5.329	2008	2700	1	4.322
##	28178	38849111746	4.527	2008	2700	1	3.344
##	28717	71749092363	3.897	2009	2700	1	2.960
##	28723	71749117733	3.826	2009	2700	1	2.721
##	28747	71549136477	3.762	2009	2700	1	2.657
##	28791	84879265065	3.850	2009	2700	1	2.913
##	29234	70450081001	5.303	2009	2700	1	4.374
##	29257	71249104673	3.549	2009	2700	1	2.612
##	29288	70449650655	4.865	2009	2700	1	3.936
##	29359	72949090601	3.696	2009	2700	1	2.591
##	29462	70350306706	3.686	2009	2700	1	2.757
##	29468	70350310101	3.801	2009	2700	1	2.872
##	29494	70350059842	3.645	2009	2700	1	2.716
##	29503	70350070124	3.427	2009	1000	2	2.665
##	29578	70349299086	3.947	2009	2700	1	3.018

##	29589	70349611684	5.029	2009	2700	1	4.100
##	29715	70350018319	5.363	2009	2700	1	4.601
##	29744	70349121347	3.487	2009	2700	1	2.725
##	29824	70349636710	3.618	2009	2700	1	2.856
##	29971	68349148204	3.480	2009	2700	1	2.551
##	30033	69449090023	3.531	2009	2700	1	2.769
##	30069	69449097272	3.961	2009	2700	1	3.032
##	30191	67651180816	4.434	2009	2700	1	3.505
##	30199	67650607983	5.537	2009	2700	1	4.608
##	30230	67650133635	3.650	2009	2700	1	2.545
##	30259	68149112409	4.126	2009	2700	1	3.021
##	30265	67649649676	4.676	2009	2700	1	3.747
##	30328	68849084141	3.947	2009	2700	1	2.842
##	30433	67649111182	4.381	2009	2700	1	3.452
##	30468	67650076181	3.734	2009	2700	1	2.797
##	30497	66649092312	3.784	2009	2700	1	2.679
##	30498	66649096725	3.506	2009	2700	1	2.577
##	30561	67349194271	3.561	2009	2700	1	2.799
##	30683	66249141245	4.629	2009	2700	1	3.867
##	30700	65549146500	3.623	2009	2700	1	2.861
##	30701	65649084159	5.220	2009	2700	1	4.115
##	30818	65349158356	3.757	2009	2700	1	2.828
##	31029	64049103218	3.775	2009	2700	1	3.013
##	31080	66849095444	3.874	2009	2700	1	2.769
##	31081	66849101464	3.493	2009	2700	1	2.564
##	31222	84859000596	3.640	2009	2700	1	2.711
##	31282	62849102089	3.467	2009	1000	2	2.538
##	31376	62549091982	3.453	2009	2700	1	2.516
##	31416	61649091385	4.600	2009	2700	1	3.671
##	31681	61749100051	3.645	2009	2700	1	2.708
##	31831	58749085209	3.937	2009	2700	1	3.008
##	31855	58149464319	3.327	2009	2700	1	2.565
##	31857	58749086000	4.008	2009	2700	1	3.079
##	31879	58149381762	3.982	2009	2700	1	3.053
##	31956	59249104722	3.922	2009	2700	1	2.817
##	32255	74949127809	3.870	2009	2700	1	2.765
##	32912	78649596720	3.725	2010	2700	1	2.586
##	33282	77958012855	3.473	2010	2700	1	2.501
##	33519	77956667533	3.560	2010	2700	1	2.764
##	33550	77956542068	3.527	2010	2700	1	2.555
##	33948	77956793539	3.509	2010	2700	1	2.713
##	34170	77954582454	3.673	2010	2700	1	2.710
##	34547	77953711938	3.886	2010	2700	1	2.914
##	34770	77953613410	3.813	2010	2700	1	2.850
##	34990	77951699612	4.010	2010	2700	1	3.047
##	35008	77951879165	4.717	2010	2700	1	3.578
##	35068	77950822827	4.845	2010	2700	1	3.882
##	35205	77949824961	4.292	2010	2700	1	3.320
##	35264	77749282912	4.078	2010	2700	1	3.115
##	35303	77749306261	4.234	2010	2700	1	3.438

##	35525	76749153747	3.436	2010	2700	1	2.640
##	35567	76249103292	3.999	2010	2700	1	3.027
##	35587	76749157063	3.860	2010	2700	1	2.897
##	35778	74849098087	3.649	2010	2700	1	2.686
##	35943	77949497387	3.679	2010	2700	1	2.540
##	37023	81555200747	3.941	2011	2700	1	2.781
##	37228	84859008091	3.976	2011	2700	1	2.983
##	37279	80054714432	4.426	2011	2700	1	3.266
##	37330	80053414769	3.623	2011	2700	1	2.638
##	37340	80052660257	4.061	2011	2700	1	2.901
##	37366	80053389356	3.851	2011	2700	1	2.691
##	37598	80052703117	4.390	2011	2700	1	3.405
##	37766	80052913150	4.505	2011	2700	1	3.512
##	37923	84857425592	3.788	2011	2700	1	2.628
##	38006	80655147954	3.804	2011	2700	1	2.819
##	38174	84859001319	3.560	2011	2700	1	2.743
##	38195	79959289318	3.793	2011	2700	1	2.976
##	38226	79958024942	3.880	2011	2700	1	2.895
##	38288	79959778973	3.405	2011	2700	1	2.588
##	38412	79955944811	3.589	2011	2700	1	2.604
##	38424	79957512933	4.392	2011	2700	1	3.407
##	38444	79956346875	4.014	2011	2700	1	3.197
##	38453	79956323623	3.988	2011	1000	2	2.828
##	38484	79955952987	3.985	2011	2700	1	3.000
##	38487	79955963939	3.644	2011	2700	1	2.659
##	38580	79958002064	3.545	2011	2700	1	2.560
##	38672	79955544813	4.917	2011	2700	1	3.932
##	38716	79953849295	3.530	2011	2700	1	2.537
##	38768	79953293021	4.330	2011	2700	1	3.170
##	38806	79955509218	3.846	2011	2700	1	3.029
##	38910	79952854226	4.158	2011	2700	1	3.173
##	38921	79952750627	4.316	2011	2700	1	3.156
##	39194	79951644557	4.908	2011	2700	1	3.748
##	39253	79551681376	3.782	2011	2700	1	2.622
##	39365	79251566761	3.950	2011	2700	1	2.790
##	39371	79251566511	3.650	2011	1000	2	2.665
##	39394	78651427507	4.543	2011	2700	1	3.558
##	39407	78650976873	3.748	2011	2700	1	2.763
##	40482	84870677378	4.205	2012	2700	1	3.018
##	40486	84870724930	4.168	2012	2700	1	3.148
##	41058	84869088051	3.548	2012	2700	1	2.528
##	41310	84868116559	3.548	2012	2700	1	2.528
##	41381	84866927095	3.757	2012	2700	1	2.737
##	41497	84866108983	3.901	2012	2700	1	2.889
##	41535	84865677850	4.010	2012	2700	1	2.998
##	41948	84864385250	3.537	2012	2700	1	2.525
##	42234	84862318037	3.695	2012	2700	1	2.683
##	42504	84862106813	3.576	2012	2700	1	2.556
##	42535	84860906122	3.765	2012	2700	1	2.578
##	42584	84860433135	4.068	2012	2700	1	3.056

```

## 42699 84860116942 3.548 2012 2700 1 2.536
## 42722 84859880855 4.194 2012 2700 1 3.182
## 42751 84859698552 3.677 2012 2700 1 2.665
## 42756 84859726718 3.810 2012 2700 1 2.623
## 42920 84859188542 3.964 2012 2700 1 2.777
## 42976 84858383393 3.794 2012 2700 1 2.774
## 43019 84857884564 3.453 2012 2700 1 2.609
## 43148 84862814741 4.189 2012 2700 1 3.169
## 43248 84857420826 4.143 2012 2700 1 2.956
## 43304 84858325834 3.502 2012 2700 1 2.658
## 43332 84856669252 4.530 2012 2700 1 3.518
## 43335 84859001282 3.740 2012 2700 1 2.553
## 43504 84857153683 3.554 2012 2700 1 2.542
## 43533 84855858670 4.001 2012 2700 1 3.157
## 43556 84859005251 3.744 2012 2700 1 2.732
## 43816 84863426446 5.368 2012 2700 1 4.181
## 43919 84864775157 3.592 2012 2700 1 2.748
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4056 -0.9292 -0.0144  0.7242  5.6083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.8638     0.0625   29.83 < 2e-16 ***
## FirstAuthorFemale1  0.1758     0.0188    9.37 < 2e-16 ***
## LastAuthorFemale1 -0.1675     0.0198   -8.44 < 2e-16 ***
## Year1997          -0.3850     0.0839   -4.59 4.5e-06 ***
## Year1998          -0.2891     0.0888   -3.26 0.00113 **
## Year1999          -0.2809     0.0836   -3.36 0.00078 ***
## Year2000           0.5418     0.1714    3.16 0.00158 **
## Year2001          -0.7271     0.0862   -8.43 < 2e-16 ***
## Year2002          -0.5222     0.0756   -6.90 5.2e-12 ***
## Year2003          -0.5795     0.0755   -7.67 1.7e-14 ***
## Year2004          -0.6296     0.0735   -8.57 < 2e-16 ***
## Year2005          -0.7252     0.0712  -10.19 < 2e-16 ***
## Year2006          -0.6678     0.0700   -9.54 < 2e-16 ***
## Year2007          -0.7896     0.0679  -11.63 < 2e-16 ***
## Year2008          -0.8564     0.0666  -12.85 < 2e-16 ***
## Year2009          -0.9346     0.0657  -14.23 < 2e-16 ***
## Year2010          -0.9004     0.0652  -13.80 < 2e-16 ***
## Year2011          -0.8793     0.0661  -13.30 < 2e-16 ***
## Year2012          -0.8522     0.0650  -13.10 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

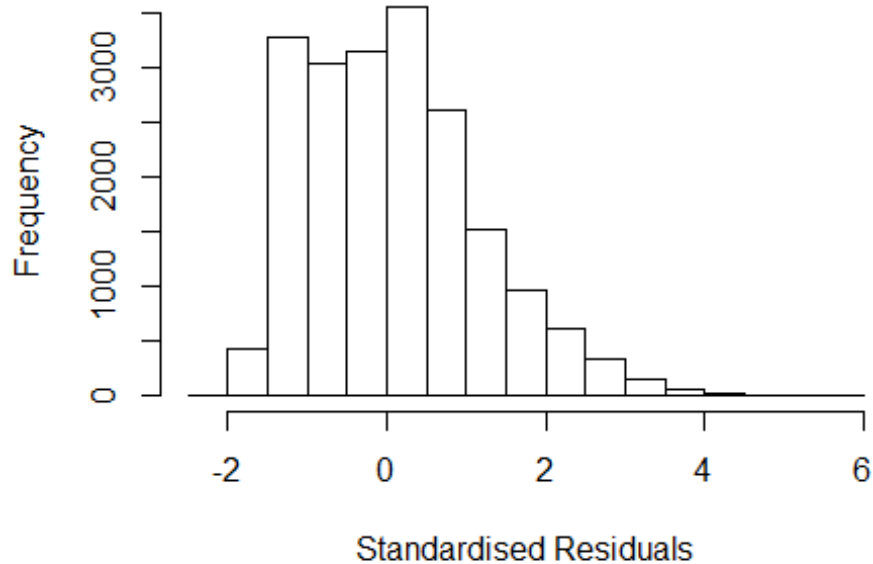
```

```

##
## Robust residual standard error: 1.06
## Multiple R-squared:  0.0524, Adjusted R-squared:  0.0515
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1639 is an outlier with |weight| = 0 ( < 5.1e-06);
## 1367 weights are ~= 1. The remaining 18346 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0013 0.8880 0.9330 0.9070 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.07e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.008 1          1.004
## Year              1.008 16          1.000

```

## Residuals from first author



```
## [1] "List of 542 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 90      0029925223 4.892 1996      2700      1      3.196
## 589     0030607603 6.092 1996      2700      1      4.396
## 613     0030599530 4.962 1996      2700      1      3.098
## 729     0029805072 5.315 1996      2700      1      3.275
## 733     0029803181 4.808 1996      2700      1      2.944
## 773     0029809469 4.580 1996      2700      1      2.716
## 852     0029785701 4.390 1996      2700      1      2.694
## 858     0029796404 4.540 1996      1300      2      2.676
## 1017    0029937218 5.035 1996      2700      1      3.339
## 1040    0030582181 4.488 1996      2700      1      2.624
## 1050    0030572468 4.501 1996      2700      1      2.637
## 1299    0030007093 4.561 1996      2700      1      2.697
## 1311    0029871557 4.455 1996      2700      1      2.583
## 1454    0029988543 4.496 1996      2700      1      2.632
## 1476    0029925181 5.956 1996      2700      1      4.092
## 1518    0030057755 4.693 1996      2700      1      2.829
## 1592    0030020685 4.440 1996      2700      1      2.576
## 1606    0030048806 4.899 1996      2700      1      2.859
## 1662    0029671048 4.411 1996      2700      1      2.547
## 1690    0030068535 4.699 1996      2700      1      3.003
## 2442    0030731968 4.336 1997      2700      1      2.857
## 2561    2642653233 4.177 1997      2700      1      2.698
## 2766    0031590608 4.218 1997      2700      1      2.907
## 2779    0030682092 4.703 1997      2700      1      3.216
## 2785    0030711655 4.498 1997      2700      1      3.019
```



##	2787	0031590552	4.556	1997	2700	1	3.077
##	2832	0031573250	5.072	1997	2700	1	3.593
##	2845	0030710196	4.233	1997	2700	1	2.922
##	2931	0030699243	4.669	1997	2700	1	3.190
##	2938	0030843904	4.847	1997	2700	1	3.192
##	2959	0030825804	4.114	1997	2700	1	2.635
##	2962	0030870938	4.251	1997	2700	1	2.940
##	2990	0030886421	3.902	1997	2700	1	2.591
##	2991	0030923557	4.251	1997	2700	1	2.596
##	3009	0030931611	4.657	1997	2700	1	3.178
##	3014	0030779141	4.141	1997	2700	1	2.662
##	3018	0030882421	4.784	1997	2700	1	3.129
##	3036	0030931456	4.323	1997	2700	1	2.836
##	3049	0030738896	4.082	1997	2700	1	2.771
##	3126	0030800984	4.242	1997	2700	1	2.763
##	3129	0030849415	4.518	1997	2700	1	3.039
##	3169	0342327375	4.622	1997	2700	1	3.143
##	3245	0030981444	4.075	1997	2700	1	2.596
##	3278	0030806335	3.985	1997	2700	1	2.506
##	3362	0031006994	4.369	1997	2700	1	2.714
##	3370	0030955988	4.114	1997	2700	1	2.635
##	3484	0342940785	4.741	1997	2700	1	3.262
##	3516	0030986151	4.459	1997	2700	1	2.980
##	3581	0008544037	5.001	1997	2700	1	3.522
##	3586	0030936556	4.929	1997	2700	1	3.274
##	3611	0030775107	4.041	1997	2700	1	2.562
##	3613	0030949484	4.274	1997	2700	1	2.795
##	3628	0030966012	4.618	1997	2700	1	3.139
##	3688	0030991825	4.401	1997	2700	1	2.922
##	3701	0030952324	4.196	1997	2700	1	2.541
##	3714	0031047332	4.445	1997	2700	1	2.790
##	3802	0031054894	4.060	1997	2700	1	2.581
##	3817	0031016791	4.320	1997	2700	1	2.841
##	3820	0031024016	3.830	1997	2700	1	2.519
##	3845	0031014243	4.141	1997	2700	1	2.662
##	3863	0031018731	4.631	1997	2700	1	3.152
##	3897	0031025486	4.377	1997	2700	1	2.898
##	4250	0030914797	4.320	1997	2700	1	2.665
##	4951	0032564695	4.550	1998	2700	1	2.975
##	5017	0032151752	4.597	1998	2700	1	3.022
##	5293	0032556176	4.149	1998	2700	1	2.742
##	5318	0032505605	4.331	1998	2700	1	2.924
##	5369	0032189103	5.551	1998	2700	1	4.144
##	5424	0032511566	6.456	1998	2700	1	4.881
##	5426	0032511583	7.183	1998	2700	1	5.608
##	5431	7344263454	4.333	1998	2700	1	2.758
##	5447	0032486979	4.238	1998	2700	1	2.663
##	5458	0032480253	4.870	1998	2700	1	3.463
##	5525	0032578283	4.135	1998	2700	1	2.560
##	5621	0032147105	4.300	1998	2700	1	2.893

## 5637	0032566160	4.969	1998	2700	1	3.394
## 5664	0032543934	4.213	1998	2700	1	2.638
## 5684	0032508294	4.995	1998	2700	1	3.588
## 5770	0032550854	4.228	1998	2700	1	2.821
## 5876	0032580764	4.527	1998	2700	1	2.952
## 5942	0032474225	4.624	1998	2700	1	3.049
## 5943	0032474229	4.290	1998	2700	1	2.715
## 6132	0032521142	4.442	1998	2700	1	2.867
## 6169	0032492429	4.972	1998	2700	1	3.397
## 6232	0032574158	4.331	1998	2700	1	2.924
## 6389	0032501528	4.831	1998	2700	1	3.256
## 6410	0032477529	4.501	1998	2700	1	2.750
## 7511	0033552391	4.335	1999	2700	1	2.752
## 7513	0033552424	4.437	1999	2700	1	2.678
## 7636	0033576162	4.142	1999	2700	1	2.559
## 7663	0033539389	4.128	1999	2700	1	2.537
## 7728	0033581376	4.083	1999	2700	1	2.500
## 7744	0033546995	5.299	1999	2700	1	3.708
## 7760	0033523401	4.771	1999	2700	1	3.180
## 7913	0033620867	4.158	1999	2700	1	2.567
## 7948	0033578120	4.184	1999	2700	1	2.601
## 8054	0000093848	5.112	1999	2700	1	3.697
## 8056	0033606486	4.150	1999	2700	1	2.567
## 8061	0033606539	4.310	1999	2700	1	2.727
## 8076	0033583970	4.325	1999	2700	1	2.566
## 8206	0033535768	4.297	1999	2700	1	2.882
## 8250	0033135998	4.757	1999	2700	1	3.342
## 8366	0033608862	4.261	1999	2700	1	2.678
## 8466	0033608466	5.159	1999	2700	1	3.744
## 8482	0033056940	4.126	1999	2700	1	2.543
## 8519	0033550718	4.179	1999	2700	1	2.596
## 8584	0033616462	4.524	1999	2700	1	2.765
## 8598	0033026308	4.715	1999	2700	1	3.132
## 8600	0033596541	4.279	1999	2700	1	2.688
## 8633	0033537299	4.467	1999	2700	1	2.884
## 8637	0033537344	4.827	1999	2700	1	3.068
## 8652	0033513984	4.145	1999	2700	1	2.730
## 8746	0032741972	4.414	1999	2700	1	2.655
## 9856	0035915319	3.646	2001	2700	1	2.509
## 9949	0035655197	4.230	2001	2700	1	3.093
## 9953	0035657149	4.504	2001	2700	1	3.359
## 9955	0035657752	4.736	2001	2700	1	3.591
## 10180	0035802719	4.271	2001	2700	1	3.134
## 10343	0035968604	3.781	2001	2700	1	2.636
## 10399	0035883865	4.405	2001	2700	1	3.268
## 10401	0035883898	4.507	2001	2700	1	3.370
## 10419	0035828388	4.911	2001	2700	1	3.774
## 10421	0035828406	4.115	2001	2700	1	2.802
## 10423	0035828439	3.858	2001	2700	1	2.545
## 10511	0035949107	4.065	2001	2700	1	2.928

##	10515	0035949122	3.764	2001	2700	1	2.627
##	10539	0035908713	3.917	2001	2700	1	2.780
##	10566	0035845259	4.459	2001	2700	1	3.490
##	10567	0035845283	4.839	2001	2700	1	3.526
##	10569	0035845310	4.139	2001	2700	1	3.002
##	10667	0035928402	4.299	2001	2700	1	3.162
##	10669	0035928414	4.026	2001	2700	1	3.057
##	10707	0035859468	3.873	2001	2700	1	2.560
##	10708	0035859488	3.903	2001	2700	1	2.766
##	10731	0035822265	4.099	2001	2700	1	2.962
##	10733	0035822274	4.398	2001	2700	1	3.261
##	10791	0035938907	3.794	2001	2700	1	2.657
##	10794	0035938954	4.242	2001	2700	1	3.105
##	10796	0035938977	4.761	2001	2700	1	3.624
##	10849	0035832508	3.881	2001	2700	1	2.568
##	10869	0035795969	5.211	2001	2700	1	4.074
##	10941	0035912582	4.041	2001	2700	1	3.072
##	10967	0035849287	4.303	2001	2700	1	3.166
##	10968	0035849309	3.630	2001	2700	1	2.661
##	10973	0035849348	3.798	2001	2700	1	2.661
##	11104	0035906301	4.056	2001	2700	1	2.919
##	11141	0035820302	3.681	2001	2700	1	2.536
##	11195	0035962071	4.328	2001	2700	1	3.191
##	11270	0035799105	4.367	2001	2700	1	3.054
##	11279	0035283198	3.917	2001	2700	1	2.780
##	11300	0035941743	4.099	2001	2700	1	2.962
##	11352	0035835406	3.975	2001	2700	1	2.838
##	11356	0035835455	4.430	2001	2700	1	3.293
##	11357	0035835456	4.299	2001	2700	1	3.162
##	11358	0035835465	3.760	2001	2700	1	2.623
##	11434	0035915614	4.165	2001	2700	1	3.020
##	11435	0035915621	4.792	2001	2700	1	3.655
##	11455	0035852473	3.982	2001	2700	1	2.669
##	11464	0035814611	4.835	2001	2700	1	3.522
##	12307	0037202686	4.381	2002	2700	1	3.039
##	12313	0037202790	5.023	2002	2700	1	3.506
##	12331	0037164375	5.152	2002	2700	1	3.978
##	12463	0037180219	3.877	2002	2700	1	2.535
##	12465	0037180235	4.153	2002	2700	1	2.811
##	12474	0037168030	4.107	2002	2700	1	2.590
##	12517	0037027198	4.571	2002	2700	1	3.221
##	12608	0037151919	4.032	2002	2700	1	2.690
##	12639	0036711316	4.157	2002	2700	1	2.815
##	12707	0037194703	3.922	2002	2700	1	2.572
##	12749	0037125595	4.180	2002	2700	1	2.663
##	12850	0037183247	4.135	2002	2700	1	2.793
##	12884	0037072062	4.256	2002	2700	1	2.739
##	13012	0037142007	4.465	2002	2700	1	3.123
##	13022	0037097683	4.661	2002	2700	1	3.319
##	13040	0037042674	4.421	2002	2700	1	3.079

##	13047	0037030659	4.332	2002	2700	1	2.815
##	13066	0036606736	4.768	2002	2700	1	3.418
##	13072	0001752768	4.604	2002	1300	2	3.262
##	13087	0037198418	4.789	2002	2700	1	3.447
##	13092	0037172392	4.111	2002	2700	1	2.769
##	13093	0037172426	4.590	2002	2700	1	3.248
##	13219	0037140185	4.238	2002	2700	1	2.721
##	13221	0037140211	4.025	2002	2700	1	2.508
##	13237	0037070811	4.227	2002	2700	1	2.877
##	13251	0037029292	4.622	2002	2700	1	3.105
##	13252	0037029304	4.335	2002	2700	1	3.161
##	13292	0037196927	3.943	2002	2700	1	2.593
##	13327	0037160919	4.208	2002	2700	1	3.034
##	13343	0037117094	4.013	2002	2700	1	2.671
##	13382	0037045846	4.216	2002	2700	1	2.874
##	13430	0037160756	4.507	2002	2700	1	3.157
##	13448	0037116793	3.905	2002	2700	1	2.731
##	13452	0037116856	4.079	2002	2700	1	2.905
##	13528	0037132846	4.730	2002	2700	1	3.388
##	14358	0345118958	4.063	2003	2700	1	2.779
##	14362	0345604401	4.243	2003	2700	1	2.959
##	14398	0242654882	4.568	2003	2700	1	3.284
##	14492	0242285681	4.228	2003	2700	1	2.944
##	14511	0142186278	4.266	2003	2700	1	2.982
##	14520	0142088505	4.874	2003	2700	1	3.590
##	14536	0141751673	3.918	2003	2700	1	2.634
##	14640	0141757403	3.972	2003	2700	1	2.512
##	14707	0041859261	5.128	2003	2700	1	3.668
##	14726	0141612000	4.003	2003	2700	1	2.719
##	14737	0142186681	4.254	2003	2700	1	2.794
##	14789	0042536436	4.076	2003	2700	1	2.616
##	14804	0042658340	4.480	2003	2700	1	3.363
##	14808	0043234290	4.784	2003	2700	1	3.500
##	14923	0038340259	4.057	2003	2700	1	2.597
##	14933	0037629268	3.791	2003	2700	1	2.507
##	14938	0038304773	4.450	2003	2700	1	3.157
##	14951	0038501052	4.215	2003	2700	1	2.931
##	15269	0037454376	3.950	2003	2700	1	2.833
##	15278	0037433669	3.825	2003	2700	1	2.532
##	15292	0037420477	3.972	2003	2700	1	2.855
##	15368	0037471825	3.966	2003	2700	1	2.506
##	15369	0037471833	4.376	2003	2700	1	3.092
##	15381	0037460811	4.023	2003	2700	1	2.739
##	15399	0037443878	4.020	2003	2700	1	2.736
##	15403	0037444242	3.641	2003	2700	1	2.524
##	15415	0037426064	4.867	2003	2700	1	3.583
##	15522	0037442694	4.176	2003	2700	1	2.883
##	15536	0037425770	4.632	2003	2700	1	3.348
##	15566	0037299629	3.641	2003	2700	1	2.524
##	15641	0037464751	3.818	2003	2700	1	2.534

##	15643	0037464756	4.600	2003	2700	1	3.140
##	15661	0037452226	3.862	2003	2700	1	2.578
##	15668	0037431695	4.060	2003	2700	1	2.767
##	15671	0037431719	3.844	2003	2700	1	2.727
##	15945	0347600946	4.318	2003	2700	1	3.034
##	16179	10944259169	3.825	2004	2700	1	2.591
##	16190	10344265934	4.562	2004	2700	1	3.328
##	16202	10644273470	3.874	2004	2700	1	2.640
##	16299	13244281317	4.541	2004	1315	2	3.307
##	16383	8544279621	4.120	2004	2700	1	2.886
##	16461	8444248351	3.914	2004	2700	1	2.680
##	16513	4844229917	3.968	2004	2700	1	2.558
##	16590	4644269881	3.881	2004	2700	1	2.647
##	16704	4344675264	4.089	2004	1704	3	2.855
##	16737	4344683381	5.510	2004	2700	1	4.100
##	16765	4043128084	3.950	2004	2700	1	2.716
##	16767	4043157677	4.539	2004	2700	1	3.305
##	16865	3242768475	4.240	2004	2700	1	2.997
##	16868	3542990935	3.776	2004	2700	1	2.542
##	16889	3242677099	4.820	2004	2700	1	3.410
##	16895	3242723235	4.333	2004	2700	1	3.266
##	16896	3242727457	3.954	2004	2700	1	2.887
##	16922	3042723720	5.214	2004	2700	1	3.980
##	17001	3042700097	5.251	2004	2700	1	4.184
##	17019	2942670131	4.495	2004	2700	1	3.261
##	17135	2442700599	4.355	2004	2700	1	2.945
##	17144	2442454953	3.773	2004	2700	1	2.539
##	17157	2342592457	3.836	2004	2700	1	2.769
##	17447	1542283597	3.988	2004	2700	1	2.745
##	17515	1542301591	3.968	2004	2700	1	2.734
##	17516	1542331537	3.852	2004	2700	1	2.618
##	17529	1342308180	4.448	2004	2700	1	3.038
##	17655	0842348097	4.802	2004	2700	1	3.559
##	18126	29544444716	4.109	2005	2700	1	2.970
##	18130	29544438304	4.009	2005	2700	1	2.695
##	18166	28844504752	4.453	2005	2700	1	3.314
##	18184	28244442441	4.424	2005	2700	1	3.285
##	18186	282444481060	3.857	2005	2700	1	2.718
##	18391	27744471516	3.860	2005	2700	1	2.546
##	18516	26244444318	4.188	2005	2700	1	3.049
##	18602	25144502474	3.767	2005	2700	1	2.628
##	18678	24044487679	3.984	2005	2700	1	2.670
##	18799	23444461703	4.567	2005	2700	1	3.420
##	19062	20444410110	3.807	2005	2700	1	2.668
##	19065	20444477045	3.829	2005	2700	1	2.690
##	19168	19744380352	4.678	2005	2700	1	3.539
##	19185	19344368841	5.322	2005	2700	1	4.351
##	19194	18744369929	4.646	2005	2700	1	3.507
##	19265	18044388395	4.220	2005	2700	1	2.906
##	19314	17244373777	3.838	2005	1000	2	2.524

##	19413	20144387123	4.557	2005	2700	1	3.410
##	19429	15244361216	5.216	2005	2700	1	4.077
##	19457	20044365983	4.316	2005	2700	1	3.177
##	19531	14644394273	4.111	2005	2700	1	2.972
##	19550	19744382704	3.823	2005	2700	1	2.684
##	19674	12344278985	3.664	2005	2700	1	2.525
##	19914	26944466840	5.136	2005	2700	1	3.997
##	19915	26944468736	4.069	2005	2700	1	2.755
##	20171	33845655265	3.983	2006	2700	1	2.787
##	20190	33845928026	4.180	2006	2700	1	2.984
##	20207	33845216421	4.276	2006	2700	1	3.072
##	20208	33845219790	3.707	2006	2700	1	2.511
##	20517	33845269095	3.712	2006	2700	1	2.516
##	20531	33751337383	4.059	2006	2700	1	2.863
##	20568	33750626215	4.171	2006	2700	1	2.967
##	20571	33750925581	4.007	2006	2700	1	2.811
##	20577	33750500333	4.675	2006	2700	1	3.303
##	20713	33749626549	3.883	2006	2700	1	2.687
##	20717	33750616076	3.876	2006	2700	1	2.504
##	20762	33750876203	3.908	2006	2700	1	2.712
##	20857	33748476238	3.824	2006	2700	1	2.628
##	20874	33748367018	4.269	2006	2700	1	3.073
##	20888	33748416499	5.205	2006	2700	1	4.176
##	20917	33748296240	4.269	2006	2700	1	2.897
##	21004	33747758047	4.019	2006	2700	1	2.647
##	21158	33746384604	4.618	2006	2700	1	3.422
##	21171	33746304156	4.304	2006	2700	1	3.100
##	21193	33745825433	4.090	2006	2700	1	2.894
##	21207	33745615112	3.747	2006	2700	1	2.551
##	21309	33745658837	4.213	2006	2700	1	2.841
##	21317	33745440706	3.989	2006	2700	1	2.793
##	21320	33745727486	3.584	2006	2700	1	2.555
##	21505	33646786398	4.166	2006	2700	1	2.970
##	21541	33646674466	4.119	2006	2700	1	2.923
##	21773	33645728134	4.034	2006	2700	1	2.838
##	21891	33644837550	3.725	2006	2700	1	2.529
##	22001	33344466383	4.144	2006	2700	1	2.948
##	22033	32644490529	4.692	2006	2700	1	3.496
##	22051	31844439769	4.084	2006	2700	1	2.888
##	22123	31444432059	3.854	2006	2700	1	2.658
##	22159	29844444176	4.173	2006	2700	1	2.977
##	22625	37049013701	5.171	2007	2700	1	4.088
##	22627	37049030806	3.471	2007	2700	1	2.564
##	22646	38049023073	4.141	2007	2700	1	3.067
##	22909	36148996837	3.610	2007	2700	1	2.703
##	22940	36348944504	5.457	2007	2700	1	4.207
##	22953	35548943929	4.812	2007	2700	1	3.562
##	22986	36849054594	3.515	2007	2700	1	2.608
##	22993	38149137150	3.935	2007	2700	1	3.028
##	23113	35548987593	3.973	2007	2700	1	2.890

##	23115	35349011537	4.544	2007	2700	1	3.470
##	23141	35248901000	4.436	2007	2700	1	3.362
##	23165	34948834261	4.114	2007	2700	1	3.207
##	23173	34548509498	4.045	2007	2700	1	3.138
##	23272	35148842531	4.087	2007	2700	1	3.180
##	23322	34548859010	3.950	2007	2700	1	2.867
##	23323	34548861255	3.930	2007	2700	1	2.680
##	23342	34548430156	3.639	2007	2700	1	2.565
##	23343	34548616859	3.747	2007	2700	1	2.664
##	23407	34548513296	3.771	2007	2700	1	2.688
##	23541	34547837402	4.813	2007	2700	1	3.739
##	23546	34548175719	4.103	2007	2700	1	3.029
##	23563	34547693119	5.506	2007	2700	1	4.599
##	23572	34547764305	4.959	2007	2700	1	3.885
##	23597	34547644454	5.013	2007	2700	1	3.939
##	23692	34547148986	3.610	2007	2700	1	2.536
##	23723	34447255724	4.649	2007	2700	1	3.575
##	23727	34447630803	4.674	2007	2700	1	3.424
##	23736	34447319145	4.410	2007	2700	1	3.336
##	23747	34447572979	4.812	2007	2700	1	3.729
##	23766	34447126847	4.160	2007	2700	1	3.086
##	23767	34447128057	3.688	2007	2700	1	2.614
##	23881	34447105600	3.765	2007	2700	1	2.515
##	23959	34250375564	4.631	2007	2700	1	3.724
##	23965	34250003794	4.837	2007	2700	1	3.587
##	24006	34347235863	3.646	2007	2700	1	2.572
##	24061	36248953716	4.289	2007	2700	1	3.039
##	24131	34249822239	3.715	2007	2700	1	2.632
##	24143	34249281687	3.954	2007	2700	1	2.880
##	24145	34249341210	4.103	2007	2700	1	3.029
##	24158	34248157118	4.876	2007	2700	1	3.802
##	24388	34047244736	3.702	2007	2700	1	2.795
##	24411	34247190031	3.702	2007	2700	1	2.628
##	24526	33847677585	4.250	2007	2700	1	3.176
##	24530	34247185441	4.298	2007	2700	1	3.215
##	24548	33947194241	4.947	2007	2700	1	3.873
##	24549	33947228816	4.850	2007	2700	1	3.943
##	24550	33947241772	4.874	2007	2700	1	3.967
##	24603	33947507103	4.263	2007	2700	1	3.180
##	24797	33847152253	4.577	2007	2700	1	3.503
##	24840	33846627856	4.213	2007	2700	1	3.139
##	24943	33846461161	4.331	2007	2700	1	3.257
##	24967	33846049673	4.834	2007	2700	1	3.760
##	24995	33846193221	3.973	2007	2700	1	2.899
##	25373	57649198083	3.905	2008	2700	1	2.898
##	25424	57749121512	4.445	2008	2700	1	3.262
##	25876	55149121818	3.553	2008	2700	1	2.713
##	25877	55149122038	5.100	2008	2700	1	4.093
##	25878	55149125165	4.787	2008	2700	1	3.780
##	25900	54949131747	4.309	2008	2700	1	3.126

##	25920	54349122201	3.593	2008	2700	1	2.586
##	25999	54849408740	3.510	2008	2700	1	2.503
##	26110	52649159199	3.642	2008	2700	1	2.626
##	26127	55849125109	4.199	2008	2700	1	3.016
##	26243	52249115660	3.717	2008	2700	1	2.710
##	26273	51049101957	3.610	2008	2700	1	2.603
##	26465	50449106919	3.416	2008	2700	1	2.576
##	26522	50149104801	3.785	2008	2700	1	2.778
##	26532	49249119141	4.376	2008	2700	1	3.369
##	26595	50949084462	3.740	2008	2700	1	2.557
##	26598	50949134384	3.348	2008	2700	1	2.508
##	26632	48249085691	3.364	2008	2700	3	2.524
##	26643	48249113707	4.179	2008	2700	1	2.996
##	26701	47149112621	5.086	2008	2700	1	4.079
##	26705	47549085541	3.510	2008	2700	1	2.503
##	26735	46949105321	4.336	2008	2700	1	3.329
##	26753	46049113677	3.962	2008	2700	1	2.779
##	26879	47149088261	4.789	2008	2700	1	3.606
##	26884	45949085378	4.364	2008	2700	1	3.357
##	27113	44849100254	3.626	2008	2700	1	2.786
##	27143	44449103216	4.014	2008	2700	1	3.174
##	27144	44449126657	3.806	2008	2700	1	2.623
##	27173	43449092540	3.839	2008	2700	1	2.832
##	27202	43749094124	3.688	2008	2700	1	2.672
##	27417	42549120016	3.688	2008	2700	1	2.848
##	27443	42249108710	3.776	2008	2700	1	2.593
##	27460	42149161021	4.109	2008	2700	1	3.102
##	27512	41849107583	4.348	2008	2700	1	3.165
##	27525	43249092442	3.668	2008	2700	1	2.661
##	27664	40949139693	4.069	2008	2700	1	3.062
##	27729	39849091058	3.927	2008	2700	1	2.920
##	27751	41549104622	3.722	2008	2700	1	2.706
##	27755	41549160713	4.051	2008	2700	1	3.044
##	27867	39649105037	4.378	2008	2700	1	3.371
##	28081	38149029249	3.437	2008	2700	1	2.597
##	28146	38049047178	5.329	2008	2700	1	4.322
##	28178	38849111746	4.527	2008	2700	1	3.344
##	28717	71749092363	3.897	2009	2700	1	2.960
##	28723	71749117733	3.826	2009	2700	1	2.721
##	28747	71549136477	3.762	2009	2700	1	2.657
##	28791	84879265065	3.850	2009	2700	1	2.913
##	29234	70450081001	5.303	2009	2700	1	4.374
##	29257	71249104673	3.549	2009	2700	1	2.612
##	29288	70449650655	4.865	2009	2700	1	3.936
##	29359	72949090601	3.696	2009	2700	1	2.591
##	29462	70350306706	3.686	2009	2700	1	2.757
##	29468	70350310101	3.801	2009	2700	1	2.872
##	29494	70350059842	3.645	2009	2700	1	2.716
##	29503	70350070124	3.427	2009	1000	2	2.665
##	29578	70349299086	3.947	2009	2700	1	3.018



##	29589	70349611684	5.029	2009	2700	1	4.100
##	29715	70350018319	5.363	2009	2700	1	4.601
##	29744	70349121347	3.487	2009	2700	1	2.725
##	29824	70349636710	3.618	2009	2700	1	2.856
##	29971	68349148204	3.480	2009	2700	1	2.551
##	30033	69449090023	3.531	2009	2700	1	2.769
##	30069	69449097272	3.961	2009	2700	1	3.032
##	30191	67651180816	4.434	2009	2700	1	3.505
##	30199	67650607983	5.537	2009	2700	1	4.608
##	30230	67650133635	3.650	2009	2700	1	2.545
##	30259	68149112409	4.126	2009	2700	1	3.021
##	30265	67649649676	4.676	2009	2700	1	3.747
##	30328	68849084141	3.947	2009	2700	1	2.842
##	30433	67649111182	4.381	2009	2700	1	3.452
##	30468	67650076181	3.734	2009	2700	1	2.797
##	30497	66649092312	3.784	2009	2700	1	2.679
##	30498	66649096725	3.506	2009	2700	1	2.577
##	30561	67349194271	3.561	2009	2700	1	2.799
##	30683	66249141245	4.629	2009	2700	1	3.867
##	30700	65549146500	3.623	2009	2700	1	2.861
##	30701	65649084159	5.220	2009	2700	1	4.115
##	30818	65349158356	3.757	2009	2700	1	2.828
##	31029	64049103218	3.775	2009	2700	1	3.013
##	31080	66849095444	3.874	2009	2700	1	2.769
##	31081	66849101464	3.493	2009	2700	1	2.564
##	31222	84859000596	3.640	2009	2700	1	2.711
##	31282	62849102089	3.467	2009	1000	2	2.538
##	31376	62549091982	3.453	2009	2700	1	2.516
##	31416	61649091385	4.600	2009	2700	1	3.671
##	31681	61749100051	3.645	2009	2700	1	2.708
##	31831	58749085209	3.937	2009	2700	1	3.008
##	31855	58149464319	3.327	2009	2700	1	2.565
##	31857	58749086000	4.008	2009	2700	1	3.079
##	31879	58149381762	3.982	2009	2700	1	3.053
##	31956	59249104722	3.922	2009	2700	1	2.817
##	32255	74949127809	3.870	2009	2700	1	2.765
##	32912	78649596720	3.725	2010	2700	1	2.586
##	33282	77958012855	3.473	2010	2700	1	2.501
##	33519	77956667533	3.560	2010	2700	1	2.764
##	33550	77956542068	3.527	2010	2700	1	2.555
##	33948	77956793539	3.509	2010	2700	1	2.713
##	34170	77954582454	3.673	2010	2700	1	2.710
##	34547	77953711938	3.886	2010	2700	1	2.914
##	34770	77953613410	3.813	2010	2700	1	2.850
##	34990	77951699612	4.010	2010	2700	1	3.047
##	35008	77951879165	4.717	2010	2700	1	3.578
##	35068	77950822827	4.845	2010	2700	1	3.882
##	35205	77949824961	4.292	2010	2700	1	3.320
##	35264	77749282912	4.078	2010	2700	1	3.115
##	35303	77749306261	4.234	2010	2700	1	3.438

##	35525	76749153747	3.436	2010	2700	1	2.640
##	35567	76249103292	3.999	2010	2700	1	3.027
##	35587	76749157063	3.860	2010	2700	1	2.897
##	35778	74849098087	3.649	2010	2700	1	2.686
##	35943	77949497387	3.679	2010	2700	1	2.540
##	37023	81555200747	3.941	2011	2700	1	2.781
##	37228	84859008091	3.976	2011	2700	1	2.983
##	37279	80054714432	4.426	2011	2700	1	3.266
##	37330	80053414769	3.623	2011	2700	1	2.638
##	37340	80052660257	4.061	2011	2700	1	2.901
##	37366	80053389356	3.851	2011	2700	1	2.691
##	37598	80052703117	4.390	2011	2700	1	3.405
##	37766	80052913150	4.505	2011	2700	1	3.512
##	37923	84857425592	3.788	2011	2700	1	2.628
##	38006	80655147954	3.804	2011	2700	1	2.819
##	38174	84859001319	3.560	2011	2700	1	2.743
##	38195	79959289318	3.793	2011	2700	1	2.976
##	38226	79958024942	3.880	2011	2700	1	2.895
##	38288	79959778973	3.405	2011	2700	1	2.588
##	38412	79955944811	3.589	2011	2700	1	2.604
##	38424	79957512933	4.392	2011	2700	1	3.407
##	38444	79956346875	4.014	2011	2700	1	3.197
##	38453	79956323623	3.988	2011	1000	2	2.828
##	38484	79955952987	3.985	2011	2700	1	3.000
##	38487	79955963939	3.644	2011	2700	1	2.659
##	38580	79958002064	3.545	2011	2700	1	2.560
##	38672	79955544813	4.917	2011	2700	1	3.932
##	38716	79953849295	3.530	2011	2700	1	2.537
##	38768	79953293021	4.330	2011	2700	1	3.170
##	38806	79955509218	3.846	2011	2700	1	3.029
##	38910	79952854226	4.158	2011	2700	1	3.173
##	38921	79952750627	4.316	2011	2700	1	3.156
##	39194	79951644557	4.908	2011	2700	1	3.748
##	39253	79551681376	3.782	2011	2700	1	2.622
##	39365	79251566761	3.950	2011	2700	1	2.790
##	39371	79251566511	3.650	2011	1000	2	2.665
##	39394	78651427507	4.543	2011	2700	1	3.558
##	39407	78650976873	3.748	2011	2700	1	2.763
##	40482	84870677378	4.205	2012	2700	1	3.018
##	40486	84870724930	4.168	2012	2700	1	3.148
##	41058	84869088051	3.548	2012	2700	1	2.528
##	41310	84868116559	3.548	2012	2700	1	2.528
##	41381	84866927095	3.757	2012	2700	1	2.737
##	41497	84866108983	3.901	2012	2700	1	2.889
##	41535	84865677850	4.010	2012	2700	1	2.998
##	41948	84864385250	3.537	2012	2700	1	2.525
##	42234	84862318037	3.695	2012	2700	1	2.683
##	42504	84862106813	3.576	2012	2700	1	2.556
##	42535	84860906122	3.765	2012	2700	1	2.578
##	42584	84860433135	4.068	2012	2700	1	3.056

```

## 42699 84860116942 3.548 2012 2700 1 2.536
## 42722 84859880855 4.194 2012 2700 1 3.182
## 42751 84859698552 3.677 2012 2700 1 2.665
## 42756 84859726718 3.810 2012 2700 1 2.623
## 42920 84859188542 3.964 2012 2700 1 2.777
## 42976 84858383393 3.794 2012 2700 1 2.774
## 43019 84857884564 3.453 2012 2700 1 2.609
## 43148 84862814741 4.189 2012 2700 1 3.169
## 43248 84857420826 4.143 2012 2700 1 2.956
## 43304 84858325834 3.502 2012 2700 1 2.658
## 43332 84856669252 4.530 2012 2700 1 3.518
## 43335 84859001282 3.740 2012 2700 1 2.553
## 43504 84857153683 3.554 2012 2700 1 2.542
## 43533 84855858670 4.001 2012 2700 1 3.157
## 43556 84859005251 3.744 2012 2700 1 2.732
## 43816 84863426446 5.368 2012 2700 1 4.181
## 43919 84864775157 3.592 2012 2700 1 2.748
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.39292 -0.93094 -0.00546  0.71430  5.61380
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.8502     0.0624   29.63 < 2e-16 ***
## FirstAuthorFemale1  0.0954     0.0161    5.94 2.9e-09 ***
## Year1997         -0.3852     0.0838   -4.60 4.3e-06 ***
## Year1998         -0.2810     0.0886   -3.17 0.00151 **
## Year1999         -0.2811     0.0836   -3.36 0.00078 ***
## Year2000          0.5427     0.1715    3.16 0.00156 **
## Year2001         -0.7385     0.0862   -8.57 < 2e-16 ***
## Year2002         -0.5257     0.0756   -6.95 3.7e-12 ***
## Year2003         -0.5872     0.0756   -7.77 8.4e-15 ***
## Year2004         -0.6395     0.0735   -8.70 < 2e-16 ***
## Year2005         -0.7330     0.0712  -10.29 < 2e-16 ***
## Year2006         -0.6743     0.0701   -9.63 < 2e-16 ***
## Year2007         -0.7989     0.0680  -11.75 < 2e-16 ***
## Year2008         -0.8664     0.0667  -13.00 < 2e-16 ***
## Year2009         -0.9435     0.0657  -14.36 < 2e-16 ***
## Year2010         -0.9058     0.0653  -13.88 < 2e-16 ***
## Year2011         -0.8848     0.0661  -13.38 < 2e-16 ***
## Year2012         -0.8580     0.0650  -13.20 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

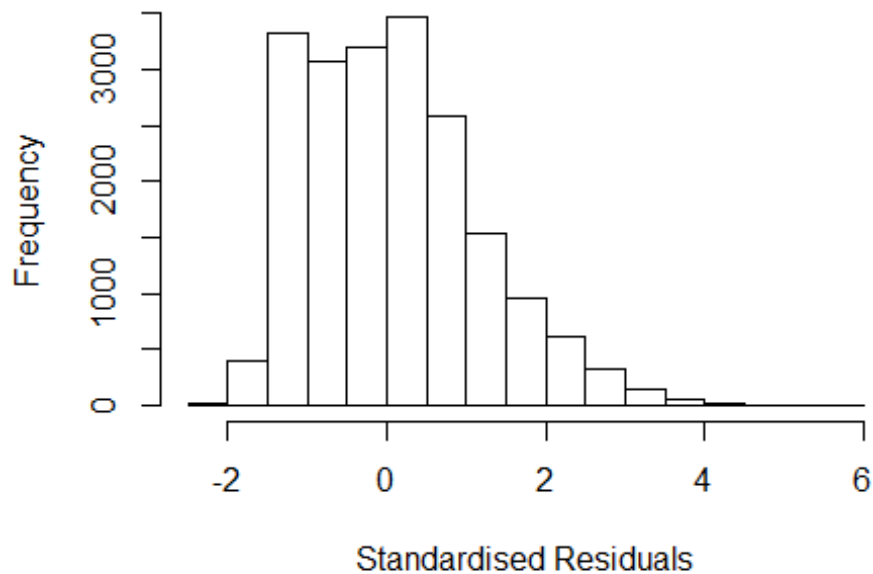
```

```

## Robust residual standard error: 1.07
## Multiple R-squared:  0.0486, Adjusted R-squared:  0.0478
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1639 is an outlier with |weight| = 0 ( < 5.1e-06);
## 1452 weights are ~= 1. The remaining 18261 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0021 0.8910 0.9350 0.9070 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.07e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.006 1          1.003
## Year            1.006 16          1.000

```

## Residuals from last author



```
## [1] "List of 542 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 90      0029925223 4.892 1996      2700      1      3.196
## 589     0030607603 6.092 1996      2700      1      4.396
## 613     0030599530 4.962 1996      2700      1      3.098
## 729     0029805072 5.315 1996      2700      1      3.275
## 733     0029803181 4.808 1996      2700      1      2.944
## 773     0029809469 4.580 1996      2700      1      2.716
## 852     0029785701 4.390 1996      2700      1      2.694
## 858     0029796404 4.540 1996      1300      2      2.676
## 1017    0029937218 5.035 1996      2700      1      3.339
## 1040    0030582181 4.488 1996      2700      1      2.624
## 1050    0030572468 4.501 1996      2700      1      2.637
## 1299    0030007093 4.561 1996      2700      1      2.697
## 1311    0029871557 4.455 1996      2700      1      2.583
## 1454    0029988543 4.496 1996      2700      1      2.632
## 1476    0029925181 5.956 1996      2700      1      4.092
## 1518    0030057755 4.693 1996      2700      1      2.829
## 1592    0030020685 4.440 1996      2700      1      2.576
## 1606    0030048806 4.899 1996      2700      1      2.859
## 1662    0029671048 4.411 1996      2700      1      2.547
## 1690    0030068535 4.699 1996      2700      1      3.003
## 2442    0030731968 4.336 1997      2700      1      2.857
## 2561    2642653233 4.177 1997      2700      1      2.698
## 2766    0031590608 4.218 1997      2700      1      2.907
## 2779    0030682092 4.703 1997      2700      1      3.216
## 2785    0030711655 4.498 1997      2700      1      3.019
```

##	2787	0031590552	4.556	1997	2700	1	3.077
##	2832	0031573250	5.072	1997	2700	1	3.593
##	2845	0030710196	4.233	1997	2700	1	2.922
##	2931	0030699243	4.669	1997	2700	1	3.190
##	2938	0030843904	4.847	1997	2700	1	3.192
##	2959	0030825804	4.114	1997	2700	1	2.635
##	2962	0030870938	4.251	1997	2700	1	2.940
##	2990	0030886421	3.902	1997	2700	1	2.591
##	2991	0030923557	4.251	1997	2700	1	2.596
##	3009	0030931611	4.657	1997	2700	1	3.178
##	3014	0030779141	4.141	1997	2700	1	2.662
##	3018	0030882421	4.784	1997	2700	1	3.129
##	3036	0030931456	4.323	1997	2700	1	2.836
##	3049	0030738896	4.082	1997	2700	1	2.771
##	3126	0030800984	4.242	1997	2700	1	2.763
##	3129	0030849415	4.518	1997	2700	1	3.039
##	3169	0342327375	4.622	1997	2700	1	3.143
##	3245	0030981444	4.075	1997	2700	1	2.596
##	3278	0030806335	3.985	1997	2700	1	2.506
##	3362	0031006994	4.369	1997	2700	1	2.714
##	3370	0030955988	4.114	1997	2700	1	2.635
##	3484	0342940785	4.741	1997	2700	1	3.262
##	3516	0030986151	4.459	1997	2700	1	2.980
##	3581	0008544037	5.001	1997	2700	1	3.522
##	3586	0030936556	4.929	1997	2700	1	3.274
##	3611	0030775107	4.041	1997	2700	1	2.562
##	3613	0030949484	4.274	1997	2700	1	2.795
##	3628	0030966012	4.618	1997	2700	1	3.139
##	3688	0030991825	4.401	1997	2700	1	2.922
##	3701	0030952324	4.196	1997	2700	1	2.541
##	3714	0031047332	4.445	1997	2700	1	2.790
##	3802	0031054894	4.060	1997	2700	1	2.581
##	3817	0031016791	4.320	1997	2700	1	2.841
##	3820	0031024016	3.830	1997	2700	1	2.519
##	3845	0031014243	4.141	1997	2700	1	2.662
##	3863	0031018731	4.631	1997	2700	1	3.152
##	3897	0031025486	4.377	1997	2700	1	2.898
##	4250	0030914797	4.320	1997	2700	1	2.665
##	4951	0032564695	4.550	1998	2700	1	2.975
##	5017	0032151752	4.597	1998	2700	1	3.022
##	5293	0032556176	4.149	1998	2700	1	2.742
##	5318	0032505605	4.331	1998	2700	1	2.924
##	5369	0032189103	5.551	1998	2700	1	4.144
##	5424	0032511566	6.456	1998	2700	1	4.881
##	5426	0032511583	7.183	1998	2700	1	5.608
##	5431	7344263454	4.333	1998	2700	1	2.758
##	5447	0032486979	4.238	1998	2700	1	2.663
##	5458	0032480253	4.870	1998	2700	1	3.463
##	5525	0032578283	4.135	1998	2700	1	2.560
##	5621	0032147105	4.300	1998	2700	1	2.893

##	5637	0032566160	4.969	1998	2700	1	3.394
##	5664	0032543934	4.213	1998	2700	1	2.638
##	5684	0032508294	4.995	1998	2700	1	3.588
##	5770	0032550854	4.228	1998	2700	1	2.821
##	5876	0032580764	4.527	1998	2700	1	2.952
##	5942	0032474225	4.624	1998	2700	1	3.049
##	5943	0032474229	4.290	1998	2700	1	2.715
##	6132	0032521142	4.442	1998	2700	1	2.867
##	6169	0032492429	4.972	1998	2700	1	3.397
##	6232	0032574158	4.331	1998	2700	1	2.924
##	6389	0032501528	4.831	1998	2700	1	3.256
##	6410	0032477529	4.501	1998	2700	1	2.750
##	7511	0033552391	4.335	1999	2700	1	2.752
##	7513	0033552424	4.437	1999	2700	1	2.678
##	7636	0033576162	4.142	1999	2700	1	2.559
##	7663	0033539389	4.128	1999	2700	1	2.537
##	7728	0033581376	4.083	1999	2700	1	2.500
##	7744	0033546995	5.299	1999	2700	1	3.708
##	7760	0033523401	4.771	1999	2700	1	3.180
##	7913	0033620867	4.158	1999	2700	1	2.567
##	7948	0033578120	4.184	1999	2700	1	2.601
##	8054	0000093848	5.112	1999	2700	1	3.697
##	8056	0033606486	4.150	1999	2700	1	2.567
##	8061	0033606539	4.310	1999	2700	1	2.727
##	8076	0033583970	4.325	1999	2700	1	2.566
##	8206	0033535768	4.297	1999	2700	1	2.882
##	8250	0033135998	4.757	1999	2700	1	3.342
##	8366	0033608862	4.261	1999	2700	1	2.678
##	8466	0033608466	5.159	1999	2700	1	3.744
##	8482	0033056940	4.126	1999	2700	1	2.543
##	8519	0033550718	4.179	1999	2700	1	2.596
##	8584	0033616462	4.524	1999	2700	1	2.765
##	8598	0033026308	4.715	1999	2700	1	3.132
##	8600	0033596541	4.279	1999	2700	1	2.688
##	8633	0033537299	4.467	1999	2700	1	2.884
##	8637	0033537344	4.827	1999	2700	1	3.068
##	8652	0033513984	4.145	1999	2700	1	2.730
##	8746	0032741972	4.414	1999	2700	1	2.655
##	9856	0035915319	3.646	2001	2700	1	2.509
##	9949	0035655197	4.230	2001	2700	1	3.093
##	9953	0035657149	4.504	2001	2700	1	3.359
##	9955	0035657752	4.736	2001	2700	1	3.591
##	10180	0035802719	4.271	2001	2700	1	3.134
##	10343	0035968604	3.781	2001	2700	1	2.636
##	10399	0035883865	4.405	2001	2700	1	3.268
##	10401	0035883898	4.507	2001	2700	1	3.370
##	10419	0035828388	4.911	2001	2700	1	3.774
##	10421	0035828406	4.115	2001	2700	1	2.802
##	10423	0035828439	3.858	2001	2700	1	2.545
##	10511	0035949107	4.065	2001	2700	1	2.928

##	10515	0035949122	3.764	2001	2700	1	2.627
##	10539	0035908713	3.917	2001	2700	1	2.780
##	10566	0035845259	4.459	2001	2700	1	3.490
##	10567	0035845283	4.839	2001	2700	1	3.526
##	10569	0035845310	4.139	2001	2700	1	3.002
##	10667	0035928402	4.299	2001	2700	1	3.162
##	10669	0035928414	4.026	2001	2700	1	3.057
##	10707	0035859468	3.873	2001	2700	1	2.560
##	10708	0035859488	3.903	2001	2700	1	2.766
##	10731	0035822265	4.099	2001	2700	1	2.962
##	10733	0035822274	4.398	2001	2700	1	3.261
##	10791	0035938907	3.794	2001	2700	1	2.657
##	10794	0035938954	4.242	2001	2700	1	3.105
##	10796	0035938977	4.761	2001	2700	1	3.624
##	10849	0035832508	3.881	2001	2700	1	2.568
##	10869	0035795969	5.211	2001	2700	1	4.074
##	10941	0035912582	4.041	2001	2700	1	3.072
##	10967	0035849287	4.303	2001	2700	1	3.166
##	10968	0035849309	3.630	2001	2700	1	2.661
##	10973	0035849348	3.798	2001	2700	1	2.661
##	11104	0035906301	4.056	2001	2700	1	2.919
##	11141	0035820302	3.681	2001	2700	1	2.536
##	11195	0035962071	4.328	2001	2700	1	3.191
##	11270	0035799105	4.367	2001	2700	1	3.054
##	11279	0035283198	3.917	2001	2700	1	2.780
##	11300	0035941743	4.099	2001	2700	1	2.962
##	11352	0035835406	3.975	2001	2700	1	2.838
##	11356	0035835455	4.430	2001	2700	1	3.293
##	11357	0035835456	4.299	2001	2700	1	3.162
##	11358	0035835465	3.760	2001	2700	1	2.623
##	11434	0035915614	4.165	2001	2700	1	3.020
##	11435	0035915621	4.792	2001	2700	1	3.655
##	11455	0035852473	3.982	2001	2700	1	2.669
##	11464	0035814611	4.835	2001	2700	1	3.522
##	12307	0037202686	4.381	2002	2700	1	3.039
##	12313	0037202790	5.023	2002	2700	1	3.506
##	12331	0037164375	5.152	2002	2700	1	3.978
##	12463	0037180219	3.877	2002	2700	1	2.535
##	12465	0037180235	4.153	2002	2700	1	2.811
##	12474	0037168030	4.107	2002	2700	1	2.590
##	12517	0037027198	4.571	2002	2700	1	3.221
##	12608	0037151919	4.032	2002	2700	1	2.690
##	12639	0036711316	4.157	2002	2700	1	2.815
##	12707	0037194703	3.922	2002	2700	1	2.572
##	12749	0037125595	4.180	2002	2700	1	2.663
##	12850	0037183247	4.135	2002	2700	1	2.793
##	12884	0037072062	4.256	2002	2700	1	2.739
##	13012	0037142007	4.465	2002	2700	1	3.123
##	13022	0037097683	4.661	2002	2700	1	3.319
##	13040	0037042674	4.421	2002	2700	1	3.079



##	13047	0037030659	4.332	2002	2700	1	2.815
##	13066	0036606736	4.768	2002	2700	1	3.418
##	13072	0001752768	4.604	2002	1300	2	3.262
##	13087	0037198418	4.789	2002	2700	1	3.447
##	13092	0037172392	4.111	2002	2700	1	2.769
##	13093	0037172426	4.590	2002	2700	1	3.248
##	13219	0037140185	4.238	2002	2700	1	2.721
##	13221	0037140211	4.025	2002	2700	1	2.508
##	13237	0037070811	4.227	2002	2700	1	2.877
##	13251	0037029292	4.622	2002	2700	1	3.105
##	13252	0037029304	4.335	2002	2700	1	3.161
##	13292	0037196927	3.943	2002	2700	1	2.593
##	13327	0037160919	4.208	2002	2700	1	3.034
##	13343	0037117094	4.013	2002	2700	1	2.671
##	13382	0037045846	4.216	2002	2700	1	2.874
##	13430	0037160756	4.507	2002	2700	1	3.157
##	13448	0037116793	3.905	2002	2700	1	2.731
##	13452	0037116856	4.079	2002	2700	1	2.905
##	13528	0037132846	4.730	2002	2700	1	3.388
##	14358	0345118958	4.063	2003	2700	1	2.779
##	14362	0345604401	4.243	2003	2700	1	2.959
##	14398	0242654882	4.568	2003	2700	1	3.284
##	14492	0242285681	4.228	2003	2700	1	2.944
##	14511	0142186278	4.266	2003	2700	1	2.982
##	14520	0142088505	4.874	2003	2700	1	3.590
##	14536	0141751673	3.918	2003	2700	1	2.634
##	14640	0141757403	3.972	2003	2700	1	2.512
##	14707	0041859261	5.128	2003	2700	1	3.668
##	14726	0141612000	4.003	2003	2700	1	2.719
##	14737	0142186681	4.254	2003	2700	1	2.794
##	14789	0042536436	4.076	2003	2700	1	2.616
##	14804	0042658340	4.480	2003	2700	1	3.363
##	14808	0043234290	4.784	2003	2700	1	3.500
##	14923	0038340259	4.057	2003	2700	1	2.597
##	14933	0037629268	3.791	2003	2700	1	2.507
##	14938	0038304773	4.450	2003	2700	1	3.157
##	14951	0038501052	4.215	2003	2700	1	2.931
##	15269	0037454376	3.950	2003	2700	1	2.833
##	15278	0037433669	3.825	2003	2700	1	2.532
##	15292	0037420477	3.972	2003	2700	1	2.855
##	15368	0037471825	3.966	2003	2700	1	2.506
##	15369	0037471833	4.376	2003	2700	1	3.092
##	15381	0037460811	4.023	2003	2700	1	2.739
##	15399	0037443878	4.020	2003	2700	1	2.736
##	15403	0037444242	3.641	2003	2700	1	2.524
##	15415	0037426064	4.867	2003	2700	1	3.583
##	15522	0037442694	4.176	2003	2700	1	2.883
##	15536	0037425770	4.632	2003	2700	1	3.348
##	15566	0037299629	3.641	2003	2700	1	2.524
##	15641	0037464751	3.818	2003	2700	1	2.534

##	15643	0037464756	4.600	2003	2700	1	3.140
##	15661	0037452226	3.862	2003	2700	1	2.578
##	15668	0037431695	4.060	2003	2700	1	2.767
##	15671	0037431719	3.844	2003	2700	1	2.727
##	15945	0347600946	4.318	2003	2700	1	3.034
##	16179	10944259169	3.825	2004	2700	1	2.591
##	16190	10344265934	4.562	2004	2700	1	3.328
##	16202	10644273470	3.874	2004	2700	1	2.640
##	16299	13244281317	4.541	2004	1315	2	3.307
##	16383	8544279621	4.120	2004	2700	1	2.886
##	16461	8444248351	3.914	2004	2700	1	2.680
##	16513	4844229917	3.968	2004	2700	1	2.558
##	16590	4644269881	3.881	2004	2700	1	2.647
##	16704	4344675264	4.089	2004	1704	3	2.855
##	16737	4344683381	5.510	2004	2700	1	4.100
##	16765	4043128084	3.950	2004	2700	1	2.716
##	16767	4043157677	4.539	2004	2700	1	3.305
##	16865	3242768475	4.240	2004	2700	1	2.997
##	16868	3542990935	3.776	2004	2700	1	2.542
##	16889	3242677099	4.820	2004	2700	1	3.410
##	16895	3242723235	4.333	2004	2700	1	3.266
##	16896	3242727457	3.954	2004	2700	1	2.887
##	16922	3042723720	5.214	2004	2700	1	3.980
##	17001	3042700097	5.251	2004	2700	1	4.184
##	17019	2942670131	4.495	2004	2700	1	3.261
##	17135	2442700599	4.355	2004	2700	1	2.945
##	17144	2442454953	3.773	2004	2700	1	2.539
##	17157	2342592457	3.836	2004	2700	1	2.769
##	17447	1542283597	3.988	2004	2700	1	2.745
##	17515	1542301591	3.968	2004	2700	1	2.734
##	17516	1542331537	3.852	2004	2700	1	2.618
##	17529	1342308180	4.448	2004	2700	1	3.038
##	17655	0842348097	4.802	2004	2700	1	3.559
##	18126	29544444716	4.109	2005	2700	1	2.970
##	18130	29544438304	4.009	2005	2700	1	2.695
##	18166	28844504752	4.453	2005	2700	1	3.314
##	18184	28244442441	4.424	2005	2700	1	3.285
##	18186	282444481060	3.857	2005	2700	1	2.718
##	18391	27744471516	3.860	2005	2700	1	2.546
##	18516	26244444318	4.188	2005	2700	1	3.049
##	18602	25144502474	3.767	2005	2700	1	2.628
##	18678	24044487679	3.984	2005	2700	1	2.670
##	18799	23444461703	4.567	2005	2700	1	3.420
##	19062	20444410110	3.807	2005	2700	1	2.668
##	19065	20444477045	3.829	2005	2700	1	2.690
##	19168	19744380352	4.678	2005	2700	1	3.539
##	19185	19344368841	5.322	2005	2700	1	4.351
##	19194	18744369929	4.646	2005	2700	1	3.507
##	19265	18044388395	4.220	2005	2700	1	2.906
##	19314	17244373777	3.838	2005	1000	2	2.524

##	19413	20144387123	4.557	2005	2700	1	3.410
##	19429	15244361216	5.216	2005	2700	1	4.077
##	19457	20044365983	4.316	2005	2700	1	3.177
##	19531	14644394273	4.111	2005	2700	1	2.972
##	19550	19744382704	3.823	2005	2700	1	2.684
##	19674	12344278985	3.664	2005	2700	1	2.525
##	19914	26944466840	5.136	2005	2700	1	3.997
##	19915	26944468736	4.069	2005	2700	1	2.755
##	20171	33845655265	3.983	2006	2700	1	2.787
##	20190	33845928026	4.180	2006	2700	1	2.984
##	20207	33845216421	4.276	2006	2700	1	3.072
##	20208	33845219790	3.707	2006	2700	1	2.511
##	20517	33845269095	3.712	2006	2700	1	2.516
##	20531	33751337383	4.059	2006	2700	1	2.863
##	20568	33750626215	4.171	2006	2700	1	2.967
##	20571	33750925581	4.007	2006	2700	1	2.811
##	20577	33750500333	4.675	2006	2700	1	3.303
##	20713	33749626549	3.883	2006	2700	1	2.687
##	20717	33750616076	3.876	2006	2700	1	2.504
##	20762	33750876203	3.908	2006	2700	1	2.712
##	20857	33748476238	3.824	2006	2700	1	2.628
##	20874	33748367018	4.269	2006	2700	1	3.073
##	20888	33748416499	5.205	2006	2700	1	4.176
##	20917	33748296240	4.269	2006	2700	1	2.897
##	21004	33747758047	4.019	2006	2700	1	2.647
##	21158	33746384604	4.618	2006	2700	1	3.422
##	21171	33746304156	4.304	2006	2700	1	3.100
##	21193	33745825433	4.090	2006	2700	1	2.894
##	21207	33745615112	3.747	2006	2700	1	2.551
##	21309	33745658837	4.213	2006	2700	1	2.841
##	21317	33745440706	3.989	2006	2700	1	2.793
##	21320	33745727486	3.584	2006	2700	1	2.555
##	21505	33646786398	4.166	2006	2700	1	2.970
##	21541	33646674466	4.119	2006	2700	1	2.923
##	21773	33645728134	4.034	2006	2700	1	2.838
##	21891	33644837550	3.725	2006	2700	1	2.529
##	22001	33344466383	4.144	2006	2700	1	2.948
##	22033	32644490529	4.692	2006	2700	1	3.496
##	22051	31844439769	4.084	2006	2700	1	2.888
##	22123	31444432059	3.854	2006	2700	1	2.658
##	22159	29844444176	4.173	2006	2700	1	2.977
##	22625	37049013701	5.171	2007	2700	1	4.088
##	22627	37049030806	3.471	2007	2700	1	2.564
##	22646	38049023073	4.141	2007	2700	1	3.067
##	22909	36148996837	3.610	2007	2700	1	2.703
##	22940	36348944504	5.457	2007	2700	1	4.207
##	22953	35548943929	4.812	2007	2700	1	3.562
##	22986	36849054594	3.515	2007	2700	1	2.608
##	22993	38149137150	3.935	2007	2700	1	3.028
##	23113	35548987593	3.973	2007	2700	1	2.890

##	23115	35349011537	4.544	2007	2700	1	3.470
##	23141	35248901000	4.436	2007	2700	1	3.362
##	23165	34948834261	4.114	2007	2700	1	3.207
##	23173	34548509498	4.045	2007	2700	1	3.138
##	23272	35148842531	4.087	2007	2700	1	3.180
##	23322	34548859010	3.950	2007	2700	1	2.867
##	23323	34548861255	3.930	2007	2700	1	2.680
##	23342	34548430156	3.639	2007	2700	1	2.565
##	23343	34548616859	3.747	2007	2700	1	2.664
##	23407	34548513296	3.771	2007	2700	1	2.688
##	23541	34547837402	4.813	2007	2700	1	3.739
##	23546	34548175719	4.103	2007	2700	1	3.029
##	23563	34547693119	5.506	2007	2700	1	4.599
##	23572	34547764305	4.959	2007	2700	1	3.885
##	23597	34547644454	5.013	2007	2700	1	3.939
##	23692	34547148986	3.610	2007	2700	1	2.536
##	23723	34447255724	4.649	2007	2700	1	3.575
##	23727	34447630803	4.674	2007	2700	1	3.424
##	23736	34447319145	4.410	2007	2700	1	3.336
##	23747	34447572979	4.812	2007	2700	1	3.729
##	23766	34447126847	4.160	2007	2700	1	3.086
##	23767	34447128057	3.688	2007	2700	1	2.614
##	23881	34447105600	3.765	2007	2700	1	2.515
##	23959	34250375564	4.631	2007	2700	1	3.724
##	23965	34250003794	4.837	2007	2700	1	3.587
##	24006	34347235863	3.646	2007	2700	1	2.572
##	24061	36248953716	4.289	2007	2700	1	3.039
##	24131	34249822239	3.715	2007	2700	1	2.632
##	24143	34249281687	3.954	2007	2700	1	2.880
##	24145	34249341210	4.103	2007	2700	1	3.029
##	24158	34248157118	4.876	2007	2700	1	3.802
##	24388	34047244736	3.702	2007	2700	1	2.795
##	24411	34247190031	3.702	2007	2700	1	2.628
##	24526	33847677585	4.250	2007	2700	1	3.176
##	24530	34247185441	4.298	2007	2700	1	3.215
##	24548	33947194241	4.947	2007	2700	1	3.873
##	24549	33947228816	4.850	2007	2700	1	3.943
##	24550	33947241772	4.874	2007	2700	1	3.967
##	24603	33947507103	4.263	2007	2700	1	3.180
##	24797	33847152253	4.577	2007	2700	1	3.503
##	24840	33846627856	4.213	2007	2700	1	3.139
##	24943	33846461161	4.331	2007	2700	1	3.257
##	24967	33846049673	4.834	2007	2700	1	3.760
##	24995	33846193221	3.973	2007	2700	1	2.899
##	25373	57649198083	3.905	2008	2700	1	2.898
##	25424	57749121512	4.445	2008	2700	1	3.262
##	25876	55149121818	3.553	2008	2700	1	2.713
##	25877	55149122038	5.100	2008	2700	1	4.093
##	25878	55149125165	4.787	2008	2700	1	3.780
##	25900	54949131747	4.309	2008	2700	1	3.126

##	25920	54349122201	3.593	2008	2700	1	2.586
##	25999	54849408740	3.510	2008	2700	1	2.503
##	26110	52649159199	3.642	2008	2700	1	2.626
##	26127	55849125109	4.199	2008	2700	1	3.016
##	26243	52249115660	3.717	2008	2700	1	2.710
##	26273	51049101957	3.610	2008	2700	1	2.603
##	26465	50449106919	3.416	2008	2700	1	2.576
##	26522	50149104801	3.785	2008	2700	1	2.778
##	26532	49249119141	4.376	2008	2700	1	3.369
##	26595	50949084462	3.740	2008	2700	1	2.557
##	26598	50949134384	3.348	2008	2700	1	2.508
##	26632	48249085691	3.364	2008	2700	3	2.524
##	26643	48249113707	4.179	2008	2700	1	2.996
##	26701	47149112621	5.086	2008	2700	1	4.079
##	26705	47549085541	3.510	2008	2700	1	2.503
##	26735	46949105321	4.336	2008	2700	1	3.329
##	26753	46049113677	3.962	2008	2700	1	2.779
##	26879	47149088261	4.789	2008	2700	1	3.606
##	26884	45949085378	4.364	2008	2700	1	3.357
##	27113	44849100254	3.626	2008	2700	1	2.786
##	27143	44449103216	4.014	2008	2700	1	3.174
##	27144	44449126657	3.806	2008	2700	1	2.623
##	27173	43449092540	3.839	2008	2700	1	2.832
##	27202	43749094124	3.688	2008	2700	1	2.672
##	27417	42549120016	3.688	2008	2700	1	2.848
##	27443	42249108710	3.776	2008	2700	1	2.593
##	27460	42149161021	4.109	2008	2700	1	3.102
##	27512	41849107583	4.348	2008	2700	1	3.165
##	27525	43249092442	3.668	2008	2700	1	2.661
##	27664	40949139693	4.069	2008	2700	1	3.062
##	27729	39849091058	3.927	2008	2700	1	2.920
##	27751	41549104622	3.722	2008	2700	1	2.706
##	27755	41549160713	4.051	2008	2700	1	3.044
##	27867	39649105037	4.378	2008	2700	1	3.371
##	28081	38149029249	3.437	2008	2700	1	2.597
##	28146	38049047178	5.329	2008	2700	1	4.322
##	28178	38849111746	4.527	2008	2700	1	3.344
##	28717	71749092363	3.897	2009	2700	1	2.960
##	28723	71749117733	3.826	2009	2700	1	2.721
##	28747	71549136477	3.762	2009	2700	1	2.657
##	28791	84879265065	3.850	2009	2700	1	2.913
##	29234	70450081001	5.303	2009	2700	1	4.374
##	29257	71249104673	3.549	2009	2700	1	2.612
##	29288	70449650655	4.865	2009	2700	1	3.936
##	29359	72949090601	3.696	2009	2700	1	2.591
##	29462	70350306706	3.686	2009	2700	1	2.757
##	29468	70350310101	3.801	2009	2700	1	2.872
##	29494	70350059842	3.645	2009	2700	1	2.716
##	29503	70350070124	3.427	2009	1000	2	2.665
##	29578	70349299086	3.947	2009	2700	1	3.018

##	29589	70349611684	5.029	2009	2700	1	4.100
##	29715	70350018319	5.363	2009	2700	1	4.601
##	29744	70349121347	3.487	2009	2700	1	2.725
##	29824	70349636710	3.618	2009	2700	1	2.856
##	29971	68349148204	3.480	2009	2700	1	2.551
##	30033	69449090023	3.531	2009	2700	1	2.769
##	30069	69449097272	3.961	2009	2700	1	3.032
##	30191	67651180816	4.434	2009	2700	1	3.505
##	30199	67650607983	5.537	2009	2700	1	4.608
##	30230	67650133635	3.650	2009	2700	1	2.545
##	30259	68149112409	4.126	2009	2700	1	3.021
##	30265	67649649676	4.676	2009	2700	1	3.747
##	30328	68849084141	3.947	2009	2700	1	2.842
##	30433	67649111182	4.381	2009	2700	1	3.452
##	30468	67650076181	3.734	2009	2700	1	2.797
##	30497	66649092312	3.784	2009	2700	1	2.679
##	30498	66649096725	3.506	2009	2700	1	2.577
##	30561	67349194271	3.561	2009	2700	1	2.799
##	30683	66249141245	4.629	2009	2700	1	3.867
##	30700	65549146500	3.623	2009	2700	1	2.861
##	30701	65649084159	5.220	2009	2700	1	4.115
##	30818	65349158356	3.757	2009	2700	1	2.828
##	31029	64049103218	3.775	2009	2700	1	3.013
##	31080	66849095444	3.874	2009	2700	1	2.769
##	31081	66849101464	3.493	2009	2700	1	2.564
##	31222	84859000596	3.640	2009	2700	1	2.711
##	31282	62849102089	3.467	2009	1000	2	2.538
##	31376	62549091982	3.453	2009	2700	1	2.516
##	31416	61649091385	4.600	2009	2700	1	3.671
##	31681	61749100051	3.645	2009	2700	1	2.708
##	31831	58749085209	3.937	2009	2700	1	3.008
##	31855	58149464319	3.327	2009	2700	1	2.565
##	31857	58749086000	4.008	2009	2700	1	3.079
##	31879	58149381762	3.982	2009	2700	1	3.053
##	31956	59249104722	3.922	2009	2700	1	2.817
##	32255	74949127809	3.870	2009	2700	1	2.765
##	32912	78649596720	3.725	2010	2700	1	2.586
##	33282	77958012855	3.473	2010	2700	1	2.501
##	33519	77956667533	3.560	2010	2700	1	2.764
##	33550	77956542068	3.527	2010	2700	1	2.555
##	33948	77956793539	3.509	2010	2700	1	2.713
##	34170	77954582454	3.673	2010	2700	1	2.710
##	34547	77953711938	3.886	2010	2700	1	2.914
##	34770	77953613410	3.813	2010	2700	1	2.850
##	34990	77951699612	4.010	2010	2700	1	3.047
##	35008	77951879165	4.717	2010	2700	1	3.578
##	35068	77950822827	4.845	2010	2700	1	3.882
##	35205	77949824961	4.292	2010	2700	1	3.320
##	35264	77749282912	4.078	2010	2700	1	3.115
##	35303	77749306261	4.234	2010	2700	1	3.438

##	35525	76749153747	3.436	2010	2700	1	2.640
##	35567	76249103292	3.999	2010	2700	1	3.027
##	35587	76749157063	3.860	2010	2700	1	2.897
##	35778	74849098087	3.649	2010	2700	1	2.686
##	35943	77949497387	3.679	2010	2700	1	2.540
##	37023	81555200747	3.941	2011	2700	1	2.781
##	37228	84859008091	3.976	2011	2700	1	2.983
##	37279	80054714432	4.426	2011	2700	1	3.266
##	37330	80053414769	3.623	2011	2700	1	2.638
##	37340	80052660257	4.061	2011	2700	1	2.901
##	37366	80053389356	3.851	2011	2700	1	2.691
##	37598	80052703117	4.390	2011	2700	1	3.405
##	37766	80052913150	4.505	2011	2700	1	3.512
##	37923	84857425592	3.788	2011	2700	1	2.628
##	38006	80655147954	3.804	2011	2700	1	2.819
##	38174	84859001319	3.560	2011	2700	1	2.743
##	38195	79959289318	3.793	2011	2700	1	2.976
##	38226	79958024942	3.880	2011	2700	1	2.895
##	38288	79959778973	3.405	2011	2700	1	2.588
##	38412	79955944811	3.589	2011	2700	1	2.604
##	38424	79957512933	4.392	2011	2700	1	3.407
##	38444	79956346875	4.014	2011	2700	1	3.197
##	38453	79956323623	3.988	2011	1000	2	2.828
##	38484	79955952987	3.985	2011	2700	1	3.000
##	38487	79955963939	3.644	2011	2700	1	2.659
##	38580	79958002064	3.545	2011	2700	1	2.560
##	38672	79955544813	4.917	2011	2700	1	3.932
##	38716	79953849295	3.530	2011	2700	1	2.537
##	38768	79953293021	4.330	2011	2700	1	3.170
##	38806	79955509218	3.846	2011	2700	1	3.029
##	38910	79952854226	4.158	2011	2700	1	3.173
##	38921	79952750627	4.316	2011	2700	1	3.156
##	39194	79951644557	4.908	2011	2700	1	3.748
##	39253	79551681376	3.782	2011	2700	1	2.622
##	39365	79251566761	3.950	2011	2700	1	2.790
##	39371	79251566511	3.650	2011	1000	2	2.665
##	39394	78651427507	4.543	2011	2700	1	3.558
##	39407	78650976873	3.748	2011	2700	1	2.763
##	40482	84870677378	4.205	2012	2700	1	3.018
##	40486	84870724930	4.168	2012	2700	1	3.148
##	41058	84869088051	3.548	2012	2700	1	2.528
##	41310	84868116559	3.548	2012	2700	1	2.528
##	41381	84866927095	3.757	2012	2700	1	2.737
##	41497	84866108983	3.901	2012	2700	1	2.889
##	41535	84865677850	4.010	2012	2700	1	2.998
##	41948	84864385250	3.537	2012	2700	1	2.525
##	42234	84862318037	3.695	2012	2700	1	2.683
##	42504	84862106813	3.576	2012	2700	1	2.556
##	42535	84860906122	3.765	2012	2700	1	2.578
##	42584	84860433135	4.068	2012	2700	1	3.056

```

## 42699 84860116942 3.548 2012 2700 1 2.536
## 42722 84859880855 4.194 2012 2700 1 3.182
## 42751 84859698552 3.677 2012 2700 1 2.665
## 42756 84859726718 3.810 2012 2700 1 2.623
## 42920 84859188542 3.964 2012 2700 1 2.777
## 42976 84858383393 3.794 2012 2700 1 2.774
## 43019 84857884564 3.453 2012 2700 1 2.609
## 43148 84862814741 4.189 2012 2700 1 3.169
## 43248 84857420826 4.143 2012 2700 1 2.956
## 43304 84858325834 3.502 2012 2700 1 2.658
## 43332 84856669252 4.530 2012 2700 1 3.518
## 43335 84859001282 3.740 2012 2700 1 2.553
## 43504 84857153683 3.554 2012 2700 1 2.542
## 43533 84855858670 4.001 2012 2700 1 3.157
## 43556 84859005251 3.744 2012 2700 1 2.732
## 43816 84863426446 5.368 2012 2700 1 4.181
## 43919 84864775157 3.592 2012 2700 1 2.748
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.4426 -0.9363 -0.0193 0.7164 5.5708
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8917 0.0627 30.18 < 2e-16 ***
## LastAuthorFemale1 -0.0740 0.0166 -4.46 8.2e-06 ***
## Year1997 -0.3811 0.0842 -4.53 6.1e-06 ***
## Year1998 -0.2795 0.0890 -3.14 0.0017 **
## Year1999 -0.2747 0.0840 -3.27 0.0011 **
## Year2000 0.5509 0.1718 3.21 0.0013 **
## Year2001 -0.7253 0.0864 -8.39 < 2e-16 ***
## Year2002 -0.5142 0.0760 -6.76 1.4e-11 ***
## Year2003 -0.5714 0.0759 -7.53 5.3e-14 ***
## Year2004 -0.6242 0.0738 -8.46 < 2e-16 ***
## Year2005 -0.7198 0.0715 -10.06 < 2e-16 ***
## Year2006 -0.6552 0.0703 -9.32 < 2e-16 ***
## Year2007 -0.7775 0.0681 -11.41 < 2e-16 ***
## Year2008 -0.8465 0.0669 -12.65 < 2e-16 ***
## Year2009 -0.9217 0.0660 -13.97 < 2e-16 ***
## Year2010 -0.8813 0.0655 -13.46 < 2e-16 ***
## Year2011 -0.8631 0.0663 -13.01 < 2e-16 ***
## Year2012 -0.8358 0.0652 -12.81 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

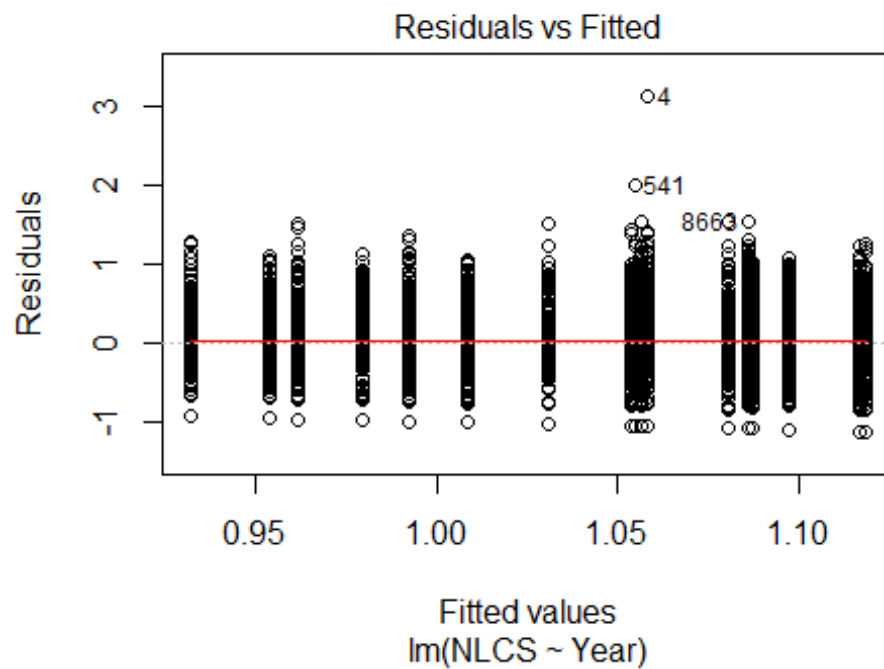


```

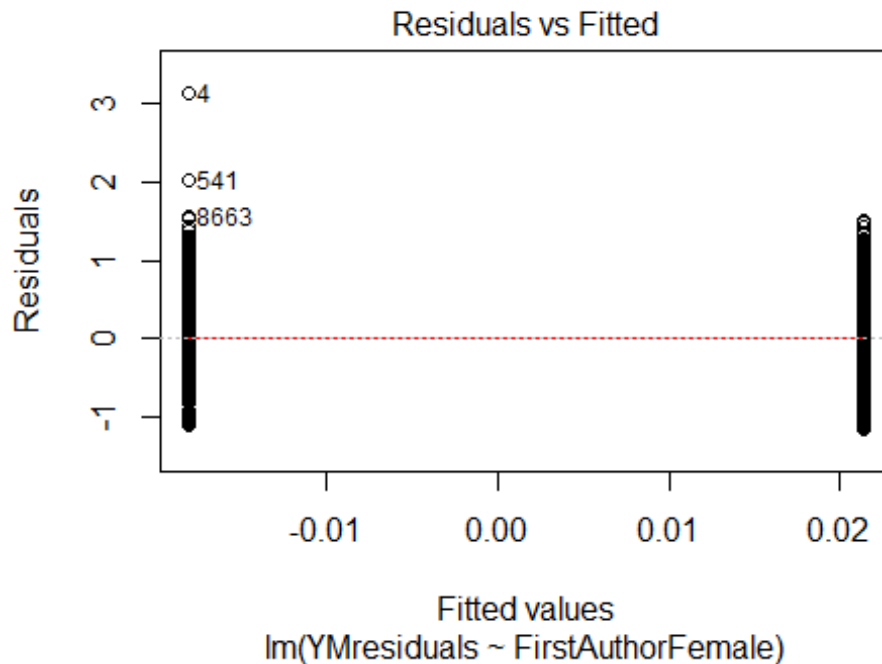
## Robust residual standard error: 1.07
## Multiple R-squared: 0.0478, Adjusted R-squared: 0.047
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1639 is an outlier with |weight| = 0 ( < 5.1e-06);
## 1437 weights are ~= 1. The remaining 18276 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0035 0.8920 0.9370 0.9070 0.9830 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.07e-06           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 19714"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2701"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 376 389 342 346 464 438 370 359 339 366 423 468 488 469 489
## 2011 2012
## 624 605
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 199 207 184 176 227 163 207 237 210 275 290 323 346 306 343
## 2011 2012
## 428 421
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 184 198 163 160 209 157 187 210 169 238 256 291 312 283 320
## 2011 2012
## 393 387
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"

```

```
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 50, df = 16, p-value = 3e-05
```

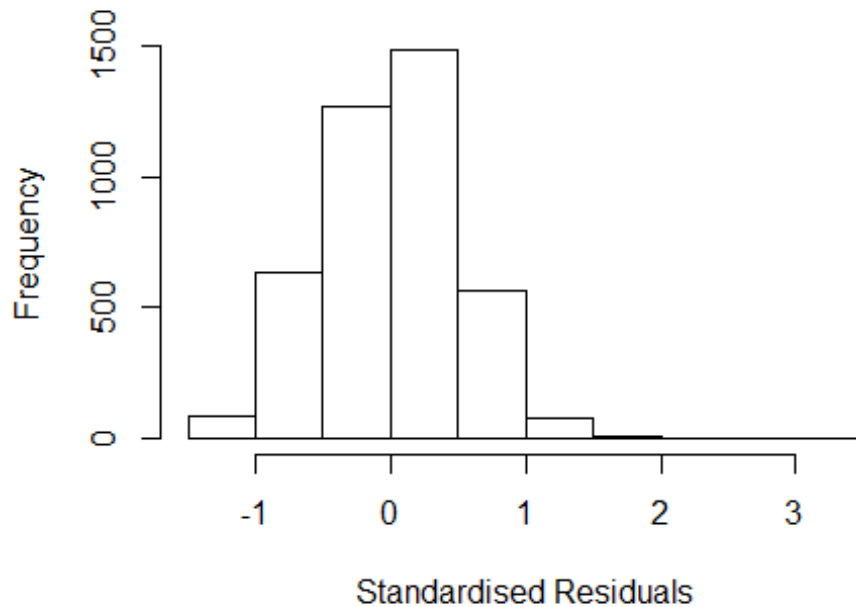


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
```



```
## [1] "Female first author team size 2018 geometric mean: 4.80247218068151"
## [1] "Male first author team size 2018 geometric mean: 3.6857437787625"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 35000, p-value = 0.001
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.40541519655056"
## [1] "Male last author team size 2018 geometric mean: 4.21613905146482"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 30000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1 1.043
## LastAuthorFemale 1.091 1 1.045
## UniqueAuthors 1.124 4 1.015
## Year 1.158 16 1.005
```

## Residuals from first and last author and team size



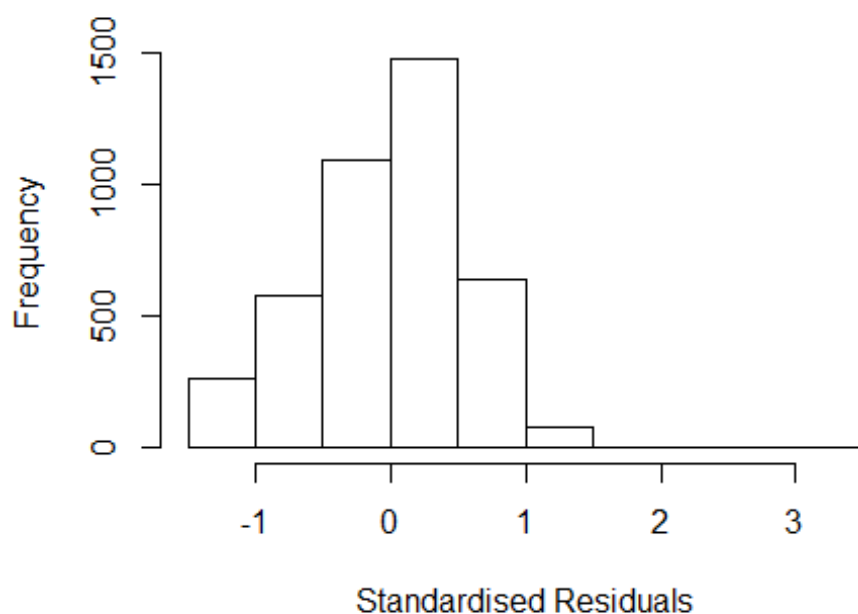
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4 0030203863 4.171 1996      2701      1      3.372
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3559 -0.3522  0.0234  0.3411  3.3718
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.799185   0.054465   14.67  <2e-16 ***
## FirstAuthorFemale1 0.014457   0.016744    0.86   0.388
## LastAuthorFemale1 -0.000957   0.017435   -0.05   0.956
## UniqueAuthors2    0.307470   0.027054   11.37  <2e-16 ***
## UniqueAuthors3    0.421725   0.025728   16.39  <2e-16 ***
## UniqueAuthors4    0.460850   0.027736   16.62  <2e-16 ***
## UniqueAuthors5    0.573959   0.022210   25.84  <2e-16 ***
## Year1997          0.014973   0.070261    0.21   0.831
## Year1998          0.061835   0.066904    0.92   0.355
## Year1999         -0.072078   0.066767   -1.08   0.280
```

```

## Year2000      -0.129295    0.065443    -1.98    0.048 *
## Year2001      -0.144939    0.067812    -2.14    0.033 *
## Year2002      -0.152598    0.065638    -2.32    0.020 *
## Year2003      -0.107969    0.061579    -1.75    0.080 .
## Year2004      -0.061860    0.063634    -0.97    0.331
## Year2005      -0.073058    0.059700    -1.22    0.221
## Year2006      -0.125722    0.059836    -2.10    0.036 *
## Year2007      -0.122547    0.060477    -2.03    0.043 *
## Year2008      -0.030717    0.059476    -0.52    0.606
## Year2009      -0.079511    0.059228    -1.34    0.180
## Year2010      -0.038990    0.060014    -0.65    0.516
## Year2011      -0.085250    0.058505    -1.46    0.145
## Year2012      -0.094693    0.058558    -1.62    0.106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.176, Adjusted R-squared:  0.172
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 3 is an outlier with |weight| = 0 ( < 2.4e-05);
## 324 weights are ~= 1. The remaining 3792 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.861  0.950  0.911  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           2.43e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample   max.it   best.r.s   k.fast.s   k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev   mts   compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
## factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1 1.043
## LastAuthorFemale 1.086 1 1.042
## Year 1.067 16 1.002

```

## Residuals from first and last author

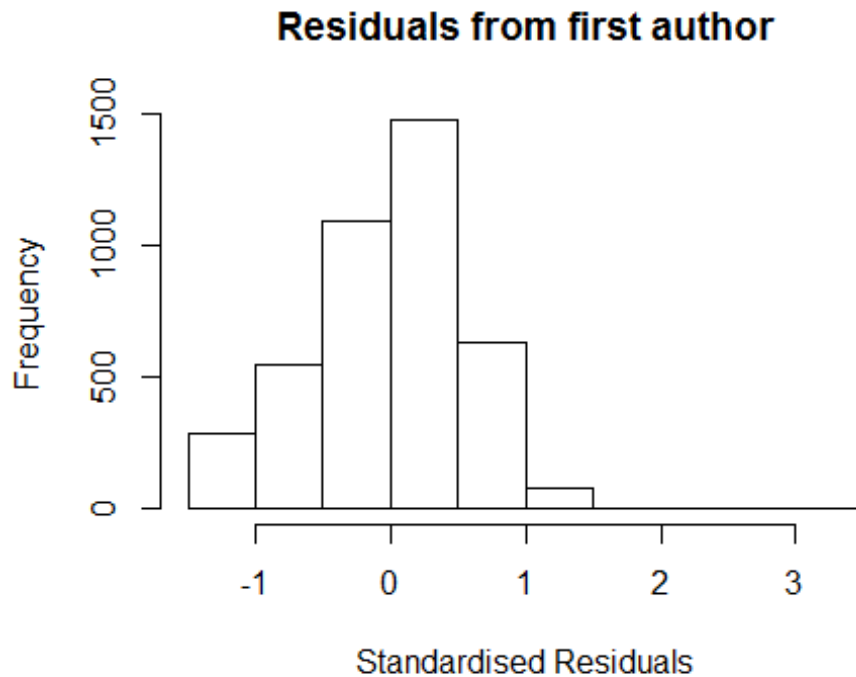


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4 0030203863 4.171 1996      2701      1      3.124
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1847 -0.3856  0.0462  0.3768  3.1243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.04668    0.05522   18.95 < 2e-16 ***
## FirstAuthorFemale1 0.06385    0.01853    3.45 0.00057 ***
## LastAuthorFemale1 -0.05963    0.01940   -3.07 0.00213 **
## Year1997          0.03471    0.07350    0.47 0.63677
## Year1998          0.08596    0.07039    1.22 0.22210
## Year1999         -0.03431    0.07245   -0.47 0.63585
## Year2000         -0.07342    0.07031   -1.04 0.29640
## Year2001         -0.11617    0.07443   -1.56 0.11862
## Year2002         -0.11238    0.07213   -1.56 0.11929
## Year2003         -0.03284    0.06732   -0.49 0.62565
## Year2004          0.04622    0.07206    0.64 0.52130
## Year2005          0.00196    0.06476    0.03 0.97587
```

```

## Year2006          -0.09644      0.06552    -1.47  0.14114
## Year2007          -0.05970      0.06456    -0.92  0.35515
## Year2008           0.07413      0.06307     1.18  0.23992
## Year2009           0.02606      0.06332     0.41  0.68068
## Year2010           0.04988      0.06332     0.79  0.43092
## Year2011           0.01224      0.06204     0.20  0.84359
## Year2012           0.03341      0.06219     0.54  0.59109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0152, Adjusted R-squared:  0.0108
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 3 is an outlier with |weight| = 0 ( < 2.4e-05);
## 320 weights are ~= 1. The remaining 3796 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.155  0.864   0.947   0.906   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042  1          1.021
## Year              1.042 16          1.001

```



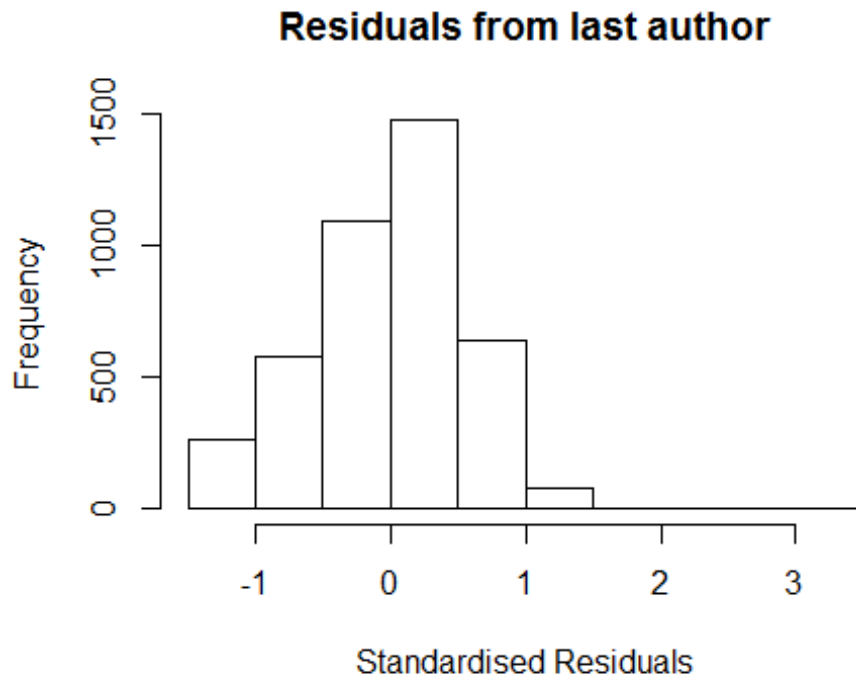
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4 0030203863 4.171 1996      2701      1      3.124
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1677 -0.3845  0.0444  0.3783  3.1342
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.036800   0.055435  18.70  <2e-16 ***
## FirstAuthorFemale1 0.044586   0.018301   2.44   0.015 *
## Year1997         0.037258   0.073676   0.51   0.613
## Year1998         0.086309   0.070825   1.22   0.223
## Year1999        -0.034525   0.072951  -0.47   0.636
## Year2000        -0.074063   0.070687  -1.05   0.295
## Year2001        -0.116205   0.074645  -1.56   0.120
## Year2002        -0.113993   0.072310  -1.58   0.115
## Year2003        -0.034164   0.067589  -0.51   0.613
## Year2004         0.047658   0.072492   0.66   0.511
## Year2005        -0.000395   0.064905  -0.01   0.995
## Year2006        -0.100272   0.065707  -1.53   0.127
```



```

## Year2007      -0.067032    0.064501   -1.04    0.299
## Year2008      0.069216    0.063147    1.10    0.273
## Year2009      0.024726    0.063579    0.39    0.697
## Year2010      0.046870    0.063498    0.74    0.460
## Year2011      0.007651    0.062105    0.12    0.902
## Year2012      0.029373    0.062235    0.47    0.637
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.013, Adjusted R-squared:  0.00893
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 3 is an outlier with |weight| = 0 ( < 2.4e-05);
## 318 weights are ~= 1. The remaining 3798 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.152  0.863   0.947   0.906   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.039 1      1.019
## Year              1.039 16      1.001

```



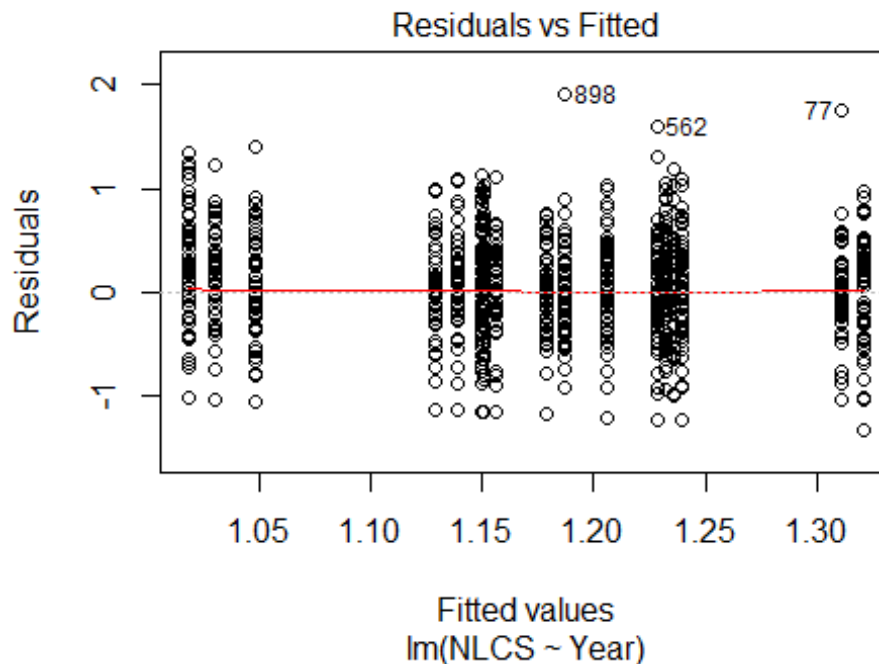
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4 0030203863 4.171 1996    2701    1    3.124
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1501 -0.3771  0.0447  0.3747  3.1108
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06016    0.05475   19.36  <2e-16 ***
## LastAuthorFemale1 -0.03723    0.01899   -1.96    0.05 *
## Year1997         0.03675    0.07297    0.50    0.61
## Year1998         0.08991    0.07032    1.28    0.20
## Year1999        -0.02564    0.07225   -0.35    0.72
## Year2000        -0.06525    0.06982   -0.93    0.35
## Year2001        -0.11052    0.07425   -1.49    0.14
## Year2002        -0.10607    0.07177   -1.48    0.14
## Year2003        -0.02565    0.06687   -0.38    0.70
## Year2004         0.05619    0.07210    0.78    0.44
## Year2005         0.00723    0.06452    0.11    0.91
## Year2006        -0.08955    0.06524   -1.37    0.17
```

```

## Year2007          -0.05110      0.06409    -0.80      0.43
## Year2008           0.08571      0.06258      1.37      0.17
## Year2009           0.03822      0.06280      0.61      0.54
## Year2010           0.06056      0.06294      0.96      0.34
## Year2011           0.02217      0.06157      0.36      0.72
## Year2012           0.04194      0.06174      0.68      0.50
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0125, Adjusted R-squared:  0.00837
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 3 is an outlier with |weight| = 0 ( < 2.4e-05);
## 328 weights are ~= 1. The remaining 3788 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.163  0.862  0.947  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4117"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2702"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 133 135 138 125 96 113 96 87 82 82 90 102 97 89 85
## 2011 2012
## 98 104
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 52 59 54 62 37 44 55 49 54 48 64 64 58 75 64

```

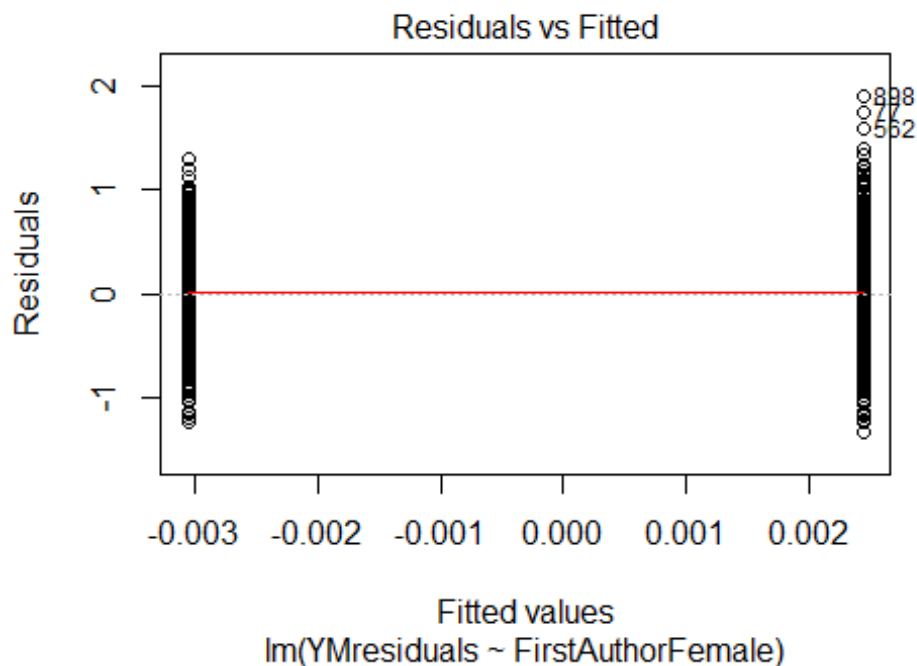
```
## 2011 2012
## 74 78
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 44 52 47 54 31 41 51 40 41 40 49 45 48 66 56
## 2011 2012
## 69 69
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.08
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 4.18833495267179"
## [1] "Male first author team size 2018 geometric mean: 4.78807406501419"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

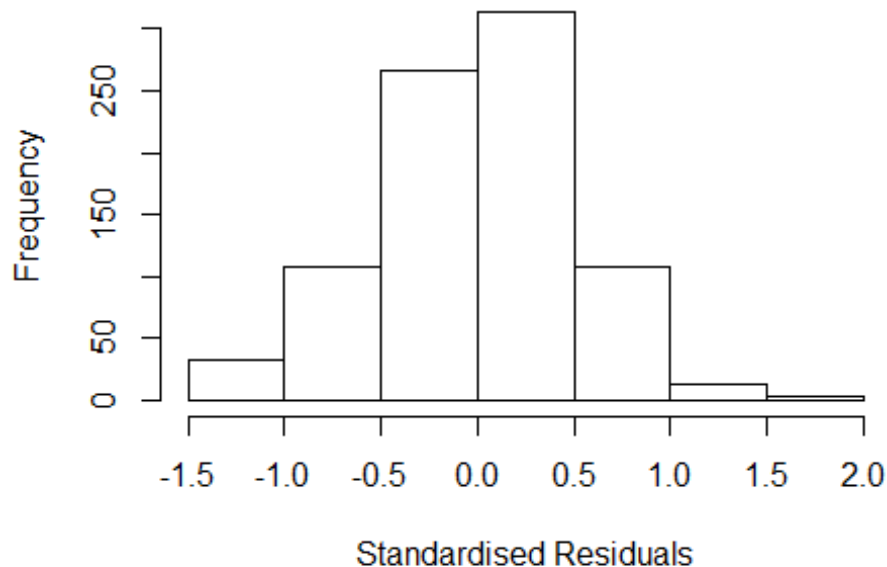
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 310, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.88185129666261"
## [1] "Male last author team size 2018 geometric mean: 4.71132831073284"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.224 1      1.106
## LastAuthorFemale  1.286 1      1.134
## UniqueAuthors    1.759 4      1.073
## Year              2.016 16     1.022
```

## Residuals from first and last author and team size



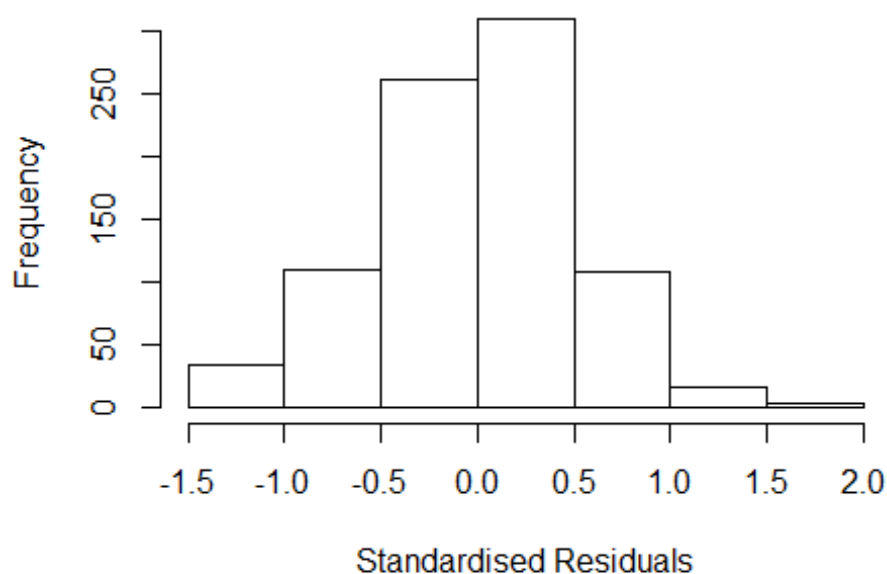
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3441 -0.3257 0.0186 0.3197 1.9828
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27091 0.07767 16.36 <2e-16 ***
## FirstAuthorFemale1 -0.00909 0.03903 -0.23 0.8159
## LastAuthorFemale1 -0.02165 0.04347 -0.50 0.6186
## UniqueAuthors2 0.01576 0.06870 0.23 0.8186
## UniqueAuthors3 0.04472 0.06429 0.70 0.4869
## UniqueAuthors4 0.02002 0.07121 0.28 0.7787
## UniqueAuthors5 0.14085 0.07077 1.99 0.0469 *
## Year1997 0.07320 0.09097 0.80 0.4212
## Year1998 -0.22822 0.11242 -2.03 0.0427 *
## Year1999 -0.08067 0.08150 -0.99 0.3226
```

```

## Year2000      -0.03283    0.10716   -0.31    0.7594
## Year2001      -0.11685    0.10930   -1.07    0.2854
## Year2002      -0.16172    0.07827   -2.07    0.0391 *
## Year2003      -0.12954    0.09827   -1.32    0.1878
## Year2004      -0.12341    0.08327   -1.48    0.1387
## Year2005      -0.17283    0.09317   -1.86    0.0639 .
## Year2006      -0.11768    0.09720   -1.21    0.2263
## Year2007       0.00134    0.09497    0.01    0.9887
## Year2008      -0.20726    0.09606   -2.16    0.0312 *
## Year2009      -0.08778    0.08563   -1.03    0.3056
## Year2010      -0.05987    0.09324   -0.64    0.5210
## Year2011      -0.11026    0.08806   -1.25    0.2109
## Year2012      -0.32313    0.10977   -2.94    0.0033 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.0435, Adjusted R-squared:  0.0179
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 761 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0634 0.8550 0.9530 0.8960 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.19e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.194 1 1.093
## LastAuthorFemale 1.243 1 1.115
## Year 1.199 16 1.006

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3914 -0.3345 0.0305 0.3176 1.9409
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29865 0.05975 21.73 <2e-16 ***
## FirstAuthorFemale1 0.00248 0.03888 0.06 0.9491
## LastAuthorFemale1 -0.03065 0.04280 -0.72 0.4741
## Year1997 0.09273 0.08821 1.05 0.2935
## Year1998 -0.20255 0.11302 -1.79 0.0735 .
## Year1999 -0.06940 0.08261 -0.84 0.4011
## Year2000 -0.00770 0.10749 -0.07 0.9429
## Year2001 -0.12524 0.10920 -1.15 0.2518
## Year2002 -0.14751 0.07952 -1.86 0.0639 .
## Year2003 -0.11466 0.09531 -1.20 0.2293
## Year2004 -0.10334 0.08132 -1.27 0.2041
## Year2005 -0.15186 0.09393 -1.62 0.1063
```

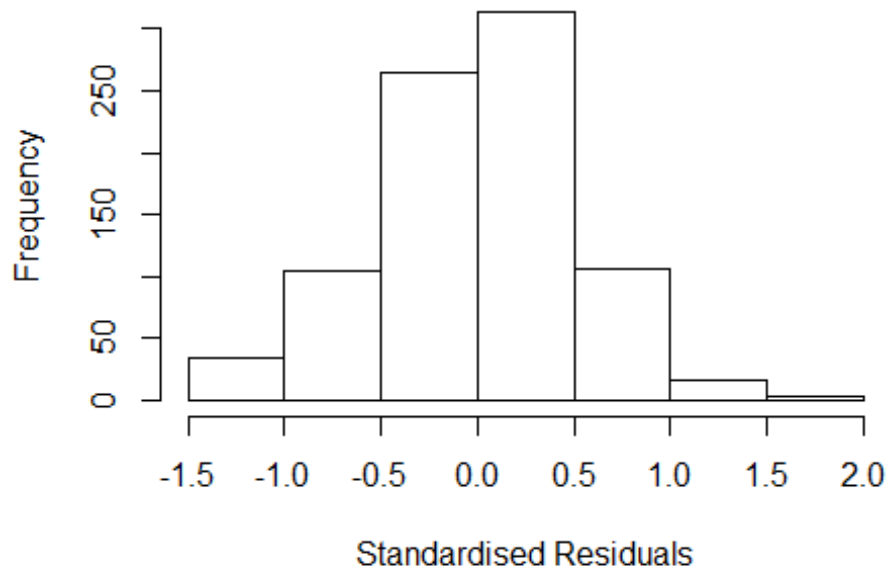


```

## Year2006      -0.08958    0.09699   -0.92    0.3560
## Year2007      0.02513    0.09547    0.26    0.7925
## Year2008     -0.17563    0.09677   -1.81    0.0699 .
## Year2009     -0.04713    0.08460   -0.56    0.5776
## Year2010     -0.03232    0.09351   -0.35    0.7297
## Year2011     -0.09497    0.08844   -1.07    0.2832
## Year2012     -0.30680    0.10941   -2.80    0.0052 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0336, Adjusted R-squared:  0.0125
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 68 weights are ~= 1. The remaining 775 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.092  0.860   0.954   0.899   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.071 1      1.035
## Year              1.071 16      1.002

```

## Residuals from first author



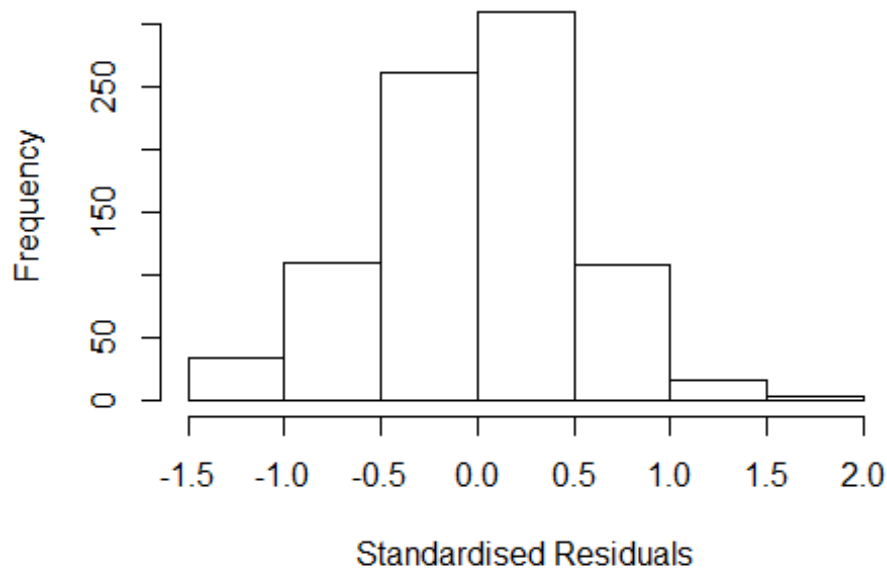
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3876 -0.3290 0.0277 0.3224 1.9452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29190 0.05888 21.94 <2e-16 ***
## FirstAuthorFemale1 -0.00607 0.03682 -0.16 0.8691
## Year1997 0.09574 0.08769 1.09 0.2752
## Year1998 -0.20045 0.11318 -1.77 0.0769 .
## Year1999 -0.07182 0.08271 -0.87 0.3854
## Year2000 -0.00197 0.10745 -0.02 0.9854
## Year2001 -0.12567 0.10946 -1.15 0.2513
## Year2002 -0.14508 0.07919 -1.83 0.0673 .
## Year2003 -0.11109 0.09464 -1.17 0.2408
## Year2004 -0.10097 0.08069 -1.25 0.2112
## Year2005 -0.14959 0.09338 -1.60 0.1095
## Year2006 -0.08756 0.09764 -0.90 0.3701
```

```

## Year2007          0.02567    0.09505    0.27    0.7872
## Year2008          -0.17501    0.09697   -1.80    0.0715 .
## Year2009          -0.05097    0.08406   -0.61    0.5445
## Year2010          -0.03039    0.09379   -0.32    0.7460
## Year2011          -0.09681    0.08797   -1.10    0.2714
## Year2012          -0.30084    0.10865   -2.77    0.0057 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0329, Adjusted R-squared:  0.013
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 72 weights are ~= 1. The remaining 771 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0889 0.8580 0.9520 0.8980 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.113 1          1.055
## Year            1.113 16          1.003

```

## Residuals from last author



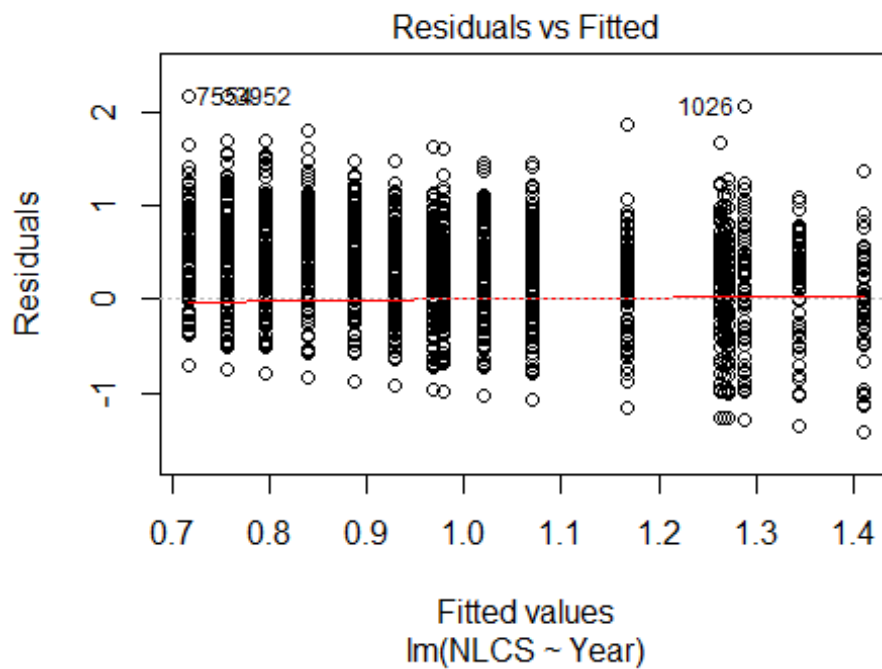
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.392 -0.334 0.030 0.317 1.940
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29947 0.05720 22.72 <2e-16 ***
## LastAuthorFemale1 -0.02982 0.04049 -0.74 0.4617
## Year1997 0.09279 0.08812 1.05 0.2927
## Year1998 -0.20230 0.11306 -1.79 0.0739 .
## Year1999 -0.06954 0.08228 -0.85 0.3983
## Year2000 -0.00767 0.10750 -0.07 0.9432
## Year2001 -0.12517 0.10915 -1.15 0.2518
## Year2002 -0.14748 0.07953 -1.85 0.0641 .
## Year2003 -0.11429 0.09523 -1.20 0.2304
## Year2004 -0.10318 0.08137 -1.27 0.2051
## Year2005 -0.15163 0.09393 -1.61 0.1069
## Year2006 -0.08969 0.09657 -0.93 0.3533
```

```

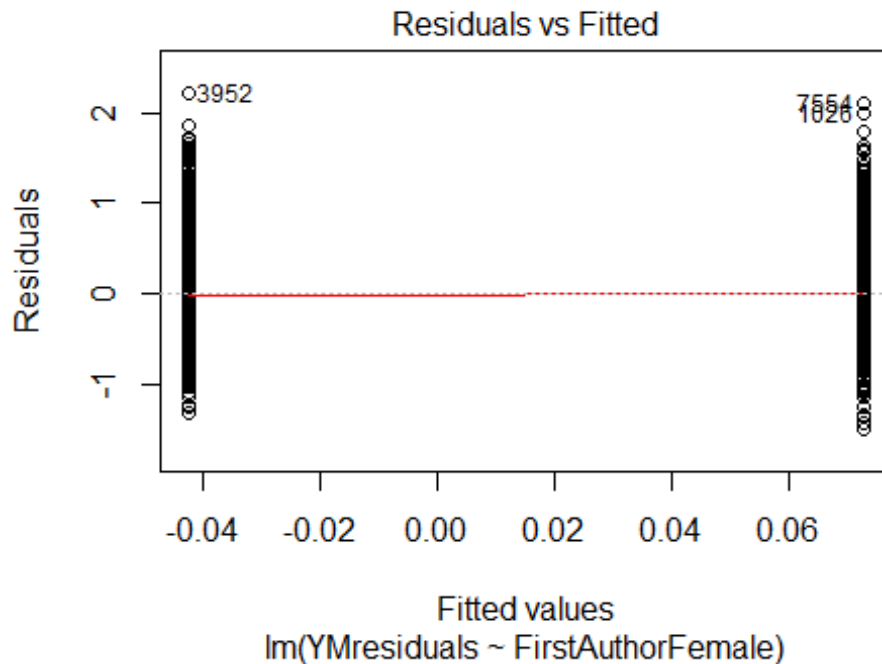
## Year2007          0.02558      0.09562      0.27      0.7891
## Year2008          -0.17561      0.09659     -1.82      0.0694 .
## Year2009          -0.04709      0.08459     -0.56      0.5779
## Year2010          -0.03225      0.09351     -0.34      0.7303
## Year2011          -0.09479      0.08856     -1.07      0.2847
## Year2012          -0.30669      0.10947     -2.80      0.0052 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.496
## Multiple R-squared:  0.0336, Adjusted R-squared:  0.0137
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 774 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.091  0.859  0.954  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 843"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2703"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 478 430 424 444 425 421 383 380 443 443 405 372 388 381 361
## 2011 2012
## 399 392
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 61 55 66 84 88 82 113 179 282 283 254 203 192 241 219
## 2011 2012

```

```
## 275 260
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 49 61 72 77 71 96 131 217 225 202 161 168 185 201
## 2011 2012
## 248 235
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```

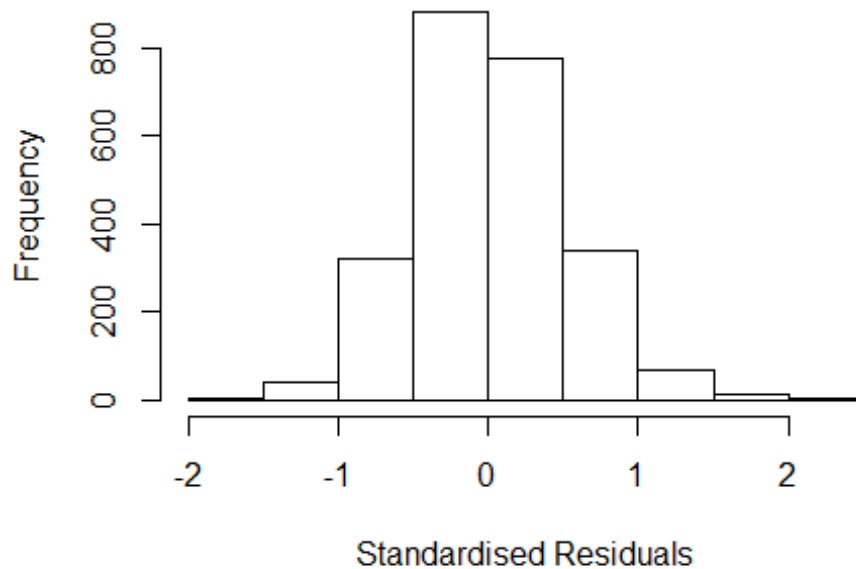


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.96, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.4626614783893"
## [1] "Male first author team size 2018 geometric mean: 2.79641340351858"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 2e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.89650222286036"
## [1] "Male last author team size 2018 geometric mean: 3.68889564431453"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.300 1          1.140
## LastAuthorFemale  1.298 1          1.139
## UniqueAuthors    1.481 4          1.050
## Year              1.509 16         1.013
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8290 -0.3445 -0.0153  0.3506  2.3054
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8842    0.0881   10.04 < 2e-16 ***
## FirstAuthorFemale1 0.0530    0.0252    2.10  0.036 *
## LastAuthorFemale1 0.0506    0.0265    1.91  0.056 .
## UniqueAuthors2    0.3894    0.0311   12.54 < 2e-16 ***
## UniqueAuthors3    0.8459    0.0379   22.35 < 2e-16 ***
## UniqueAuthors4    0.9421    0.0428   21.99 < 2e-16 ***
## UniqueAuthors5    1.0982    0.0332   33.10 < 2e-16 ***
## Year1997         -0.1534    0.1378   -1.11  0.266
## Year1998         -0.3246    0.1368   -2.37  0.018 *
## Year1999         -0.2131    0.1085   -1.96  0.050 *
```

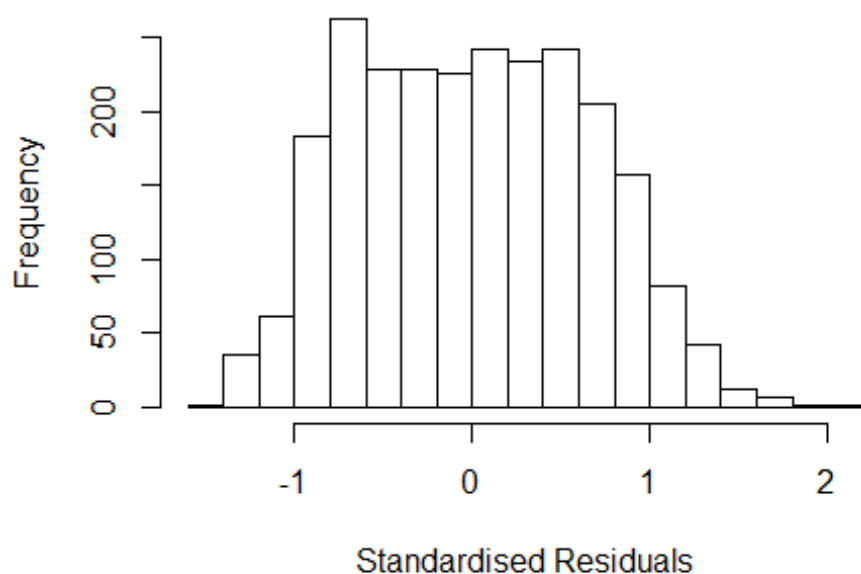


```

## Year2000      -0.2936      0.1147      -2.56      0.011 *
## Year2001      -0.2996      0.1089      -2.75      0.006 **
## Year2002      -0.4472      0.1018      -4.39      1.2e-05 ***
## Year2003      -0.5362      0.0955      -5.62      2.2e-08 ***
## Year2004      -0.6285      0.0906      -6.94      5.1e-12 ***
## Year2005      -0.5797      0.0904      -6.41      1.7e-10 ***
## Year2006      -0.4788      0.0918      -5.22      2.0e-07 ***
## Year2007      -0.4460      0.0936      -4.76      2.0e-06 ***
## Year2008      -0.5135      0.0939      -5.47      5.0e-08 ***
## Year2009      -0.5328      0.0951      -5.60      2.4e-08 ***
## Year2010      -0.5398      0.0940      -5.74      1.1e-08 ***
## Year2011      -0.6218      0.0914      -6.80      1.3e-11 ***
## Year2012      -0.6553      0.0910      -7.20      8.1e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.453, Adjusted R-squared:  0.448
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 116 is an outlier with |weight| = 0 ( < 4.1e-05);
## 192 weights are ~= 1. The remaining 2260 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0123 0.8490 0.9480 0.8930 0.9810 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.08e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev mts compute.rd
##      0      1000      0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.187 1 1.089
## LastAuthorFemale 1.185 1 1.089
## Year 1.053 16 1.002

```

## Residuals from first and last author



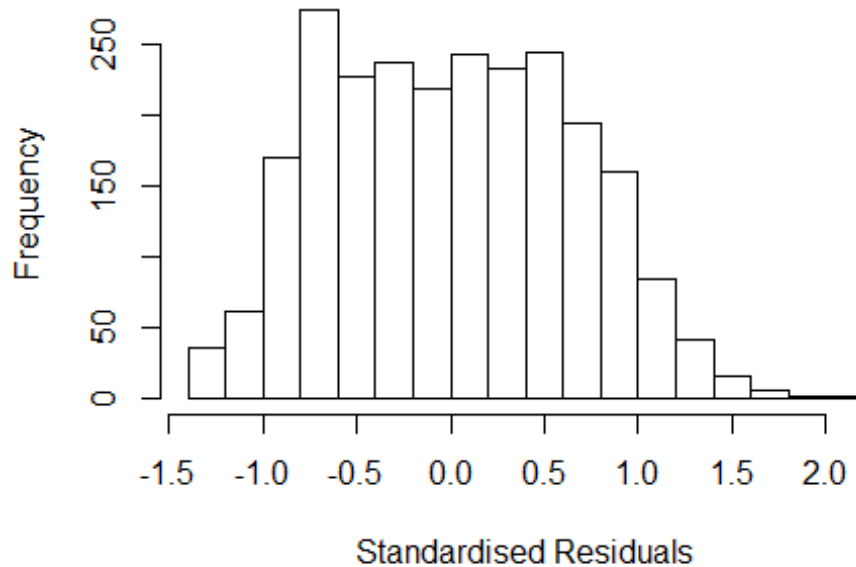
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.402624 -0.552182 -0.000282 0.504951 2.079024
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4162 0.0694 20.40 < 2e-16 ***
## FirstAuthorFemale1 0.1348 0.0318 4.24 2.3e-05 ***
## LastAuthorFemale1 0.0268 0.0348 0.77 0.44083
## Year1997 -0.1006 0.1265 -0.80 0.42649
## Year1998 -0.1752 0.1286 -1.36 0.17323
## Year1999 -0.1977 0.1160 -1.71 0.08830 .
## Year2000 -0.1844 0.0993 -1.86 0.06354 .
## Year2001 -0.2120 0.1007 -2.11 0.03534 *
## Year2002 -0.3421 0.0921 -3.71 0.00021 ***
## Year2003 -0.5947 0.0931 -6.39 2.0e-10 ***
## Year2004 -0.8641 0.0822 -10.51 < 2e-16 ***
## Year2005 -0.7600 0.0820 -9.27 < 2e-16 ***
```

```

## Year2006          -0.6433      0.0843   -7.63  3.4e-14 ***
## Year2007          -0.4139      0.0852   -4.86  1.3e-06 ***
## Year2008          -0.5386      0.0855   -6.30  3.5e-10 ***
## Year2009          -0.5055      0.0857   -5.90  4.2e-09 ***
## Year2010          -0.6117      0.0877   -6.97  4.0e-12 ***
## Year2011          -0.7233      0.0810   -8.93  < 2e-16 ***
## Year2012          -0.7739      0.0822   -9.41  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.693
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.103
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 2292 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.349  0.881  0.943  0.920  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1          1.011
## Year              1.023 16          1.001

```

## Residuals from first author



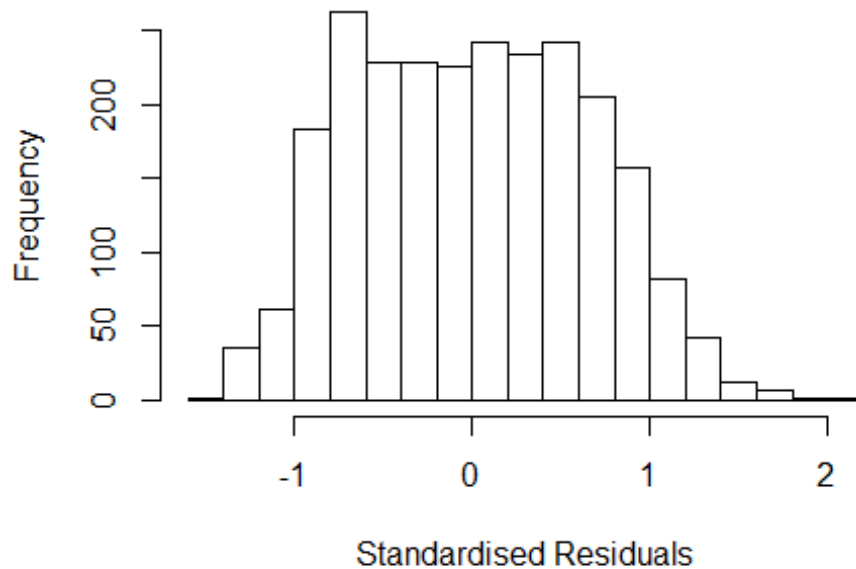
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.389479 -0.554804 -0.000638 0.505730 2.092728
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4174 0.0696 20.38 < 2e-16 ***
## FirstAuthorFemale1 0.1450 0.0295 4.91 9.5e-07 ***
## Year1997 -0.0971 0.1266 -0.77 0.44309
## Year1998 -0.1729 0.1285 -1.35 0.17874
## Year1999 -0.1967 0.1161 -1.69 0.09027 .
## Year2000 -0.1810 0.0992 -1.82 0.06827 .
## Year2001 -0.2084 0.1009 -2.07 0.03897 *
## Year2002 -0.3396 0.0922 -3.68 0.00023 ***
## Year2003 -0.5926 0.0932 -6.36 2.5e-10 ***
## Year2004 -0.8626 0.0823 -10.48 < 2e-16 ***
## Year2005 -0.7569 0.0821 -9.21 < 2e-16 ***
## Year2006 -0.6397 0.0844 -7.58 4.8e-14 ***
```

```

## Year2007          -0.4127      0.0852   -4.84  1.4e-06 ***
## Year2008          -0.5362      0.0856   -6.26  4.4e-10 ***
## Year2009          -0.5011      0.0854   -5.87  5.1e-09 ***
## Year2010          -0.6070      0.0877   -6.92  5.7e-12 ***
## Year2011          -0.7201      0.0810   -8.89  < 2e-16 ***
## Year2012          -0.7721      0.0823   -9.39  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.694
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.103
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 166 weights are ~= 1. The remaining 2287 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.342  0.881  0.943  0.920  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.010
## Year              1.021 16      1.001

```

## Residuals from last author



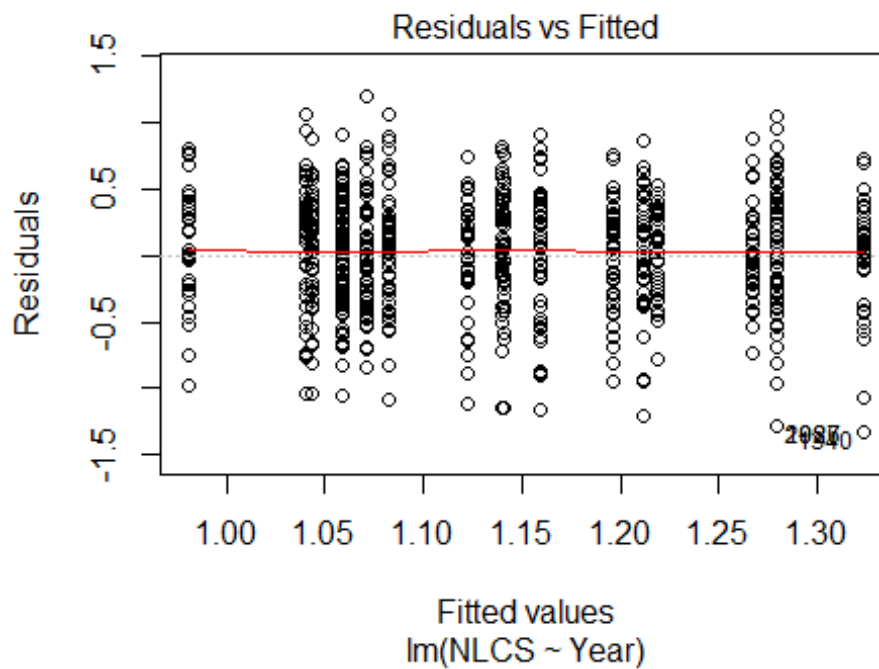
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3601 -0.5799 -0.0119 0.5128 2.1222
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4623 0.0662 22.10 < 2e-16 ***
## LastAuthorFemale1 0.0852 0.0318 2.68 0.0075 **
## Year1997 -0.1160 0.1251 -0.93 0.3540
## Year1998 -0.1874 0.1308 -1.43 0.1520
## Year1999 -0.2180 0.1153 -1.89 0.0587 .
## Year2000 -0.2020 0.0970 -2.08 0.0373 *
## Year2001 -0.2255 0.0988 -2.28 0.0225 *
## Year2002 -0.3583 0.0906 -3.96 7.8e-05 ***
## Year2003 -0.5972 0.0912 -6.55 7.1e-11 ***
## Year2004 -0.8825 0.0810 -10.90 < 2e-16 ***
## Year2005 -0.7849 0.0803 -9.78 < 2e-16 ***
## Year2006 -0.6571 0.0825 -7.96 2.6e-15 ***
```

```

## Year2007          -0.4150      0.0837   -4.96  7.6e-07 ***
## Year2008          -0.5487      0.0835   -6.57  6.1e-11 ***
## Year2009          -0.5193      0.0841   -6.17  7.8e-10 ***
## Year2010          -0.6277      0.0866   -7.25  5.6e-13 ***
## Year2011          -0.7376      0.0797   -9.26  < 2e-16 ***
## Year2012          -0.7867      0.0811   -9.70  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.691
## Multiple R-squared:  0.103, Adjusted R-squared:  0.0971
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 153 weights are ~ = 1. The remaining 2300 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.326  0.880  0.937  0.919  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.08e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2453"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2704"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   90   98   90   97   98  116  118   80   80   91   60   79   91   83   77
## 2011 2012
##   91   90
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   47   40   37   29   40   66   43   49   59   37   38   55   52   46
## 2011 2012

```

```
## 50 62
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 39 37 33 23 35 49 28 39 45 33 31 42 41 34
## 2011 2012
## 36 46
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.04
```

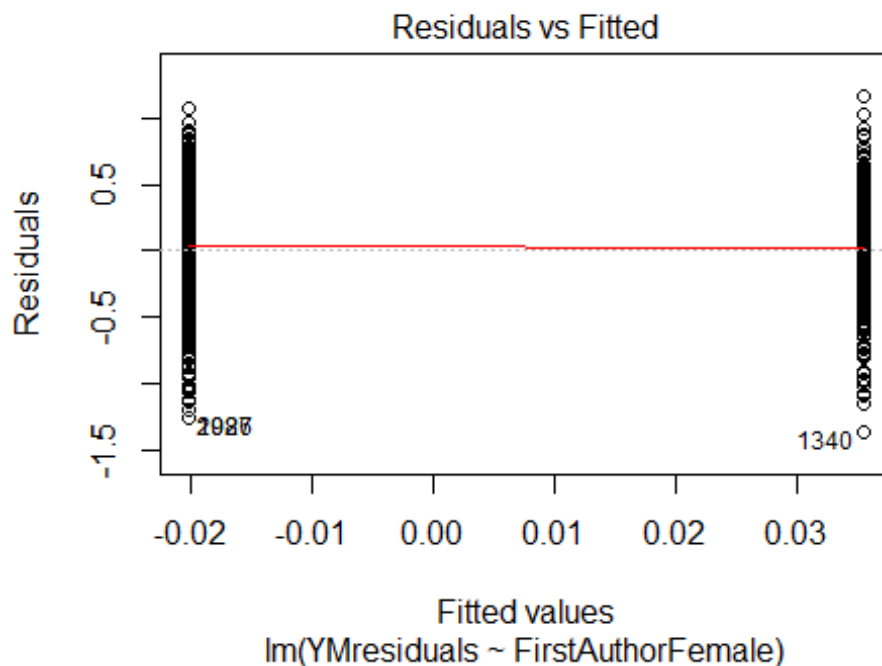


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.5, df = 1, p-value = 0.03
## [1] "Female first author team size 2018 geometric mean: 6.97195503568772"
## [1] "Male first author team size 2018 geometric mean: 3.97201112885289"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 7.33111434288597"
## [1] "Male last author team size 2018 geometric mean: 4.66693094093684"

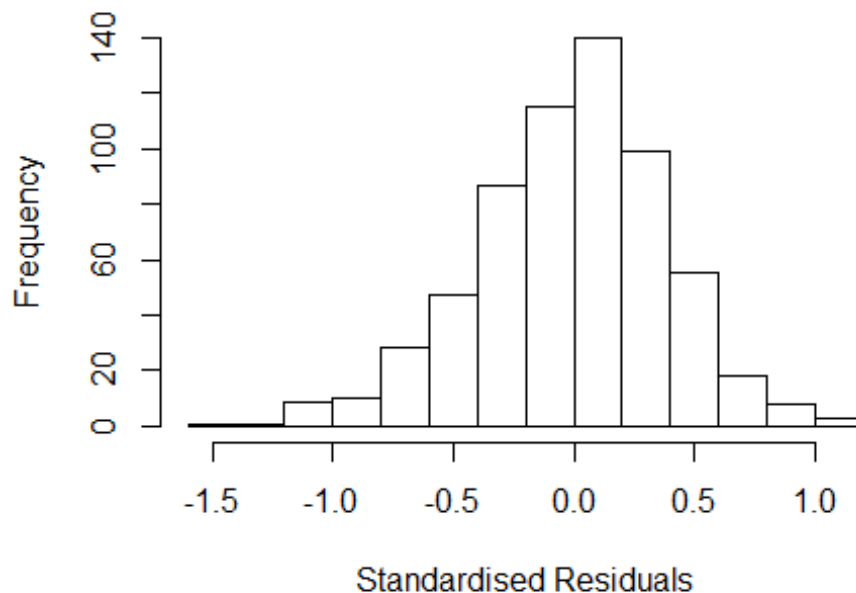
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 150, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.247	1	1.117
LastAuthorFemale	1.153	1	1.074
UniqueAuthors	1.921	4	1.085
Year	2.381	16	1.027

## Residuals from first and last author and team size



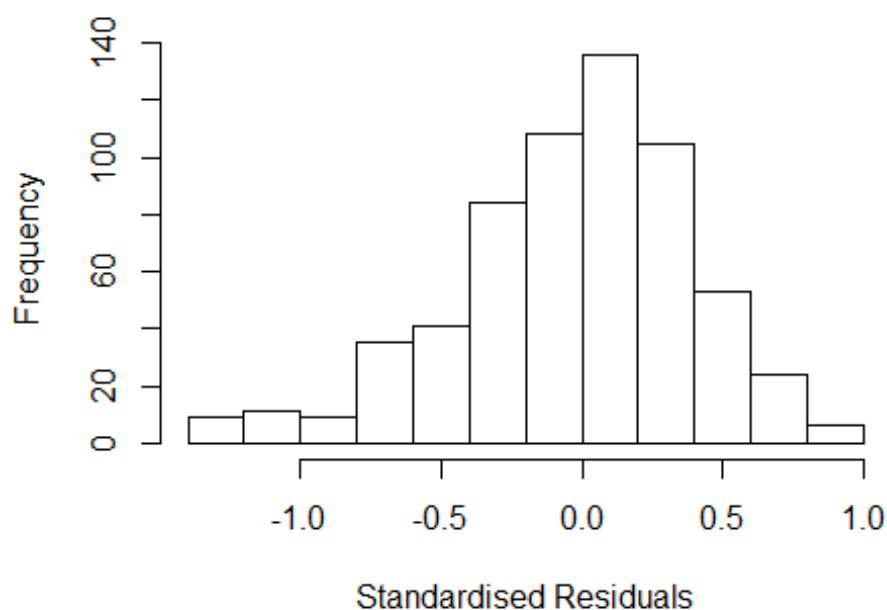
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4560 -0.2628  0.0186  0.2563  1.1664
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8908    0.0956   9.32 < 2e-16 ***
## FirstAuthorFemale1 -0.0317    0.0340  -0.93  0.35144
## LastAuthorFemale1 -0.0040    0.0405  -0.10  0.92129
## UniqueAuthors2     0.3449    0.1132   3.05  0.00242 **
## UniqueAuthors3     0.3893    0.1003   3.88  0.00012 ***
## UniqueAuthors4     0.4554    0.0977   4.66  3.9e-06 ***
## UniqueAuthors5     0.5476    0.0895   6.12  1.7e-09 ***
## Year1997         -0.0905    0.0876  -1.03  0.30192
## Year1998         -0.0848    0.0786  -1.08  0.28113
## Year1999         -0.2353    0.0891  -2.64  0.00850 **
```

```

## Year2000          -0.1429      0.1114    -1.28   0.20012
## Year2001          -0.0545      0.0831    -0.66   0.51227
## Year2002          -0.2361      0.0809    -2.92   0.00364 **
## Year2003          -0.2717      0.1092    -2.49   0.01314 *
## Year2004          -0.2626      0.0835    -3.14   0.00175 **
## Year2005          -0.2208      0.0840    -2.63   0.00881 **
## Year2006          -0.1607      0.0785    -2.05   0.04119 *
## Year2007           0.0533      0.0915     0.58   0.56050
## Year2008          -0.1516      0.0799    -1.90   0.05826 .
## Year2009          -0.0284      0.0936    -0.30   0.76209
## Year2010          -0.0885      0.0918    -0.96   0.33528
## Year2011          -0.1603      0.0910    -1.76   0.07856 .
## Year2012          -0.0702      0.0968    -0.73   0.46852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.167, Adjusted R-squared:  0.137
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 566 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0934 0.8690 0.9490 0.8980 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.61e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.148 1 1.072
## LastAuthorFemale 1.116 1 1.056
## Year 1.278 16 1.008

```

## Residuals from first and last author



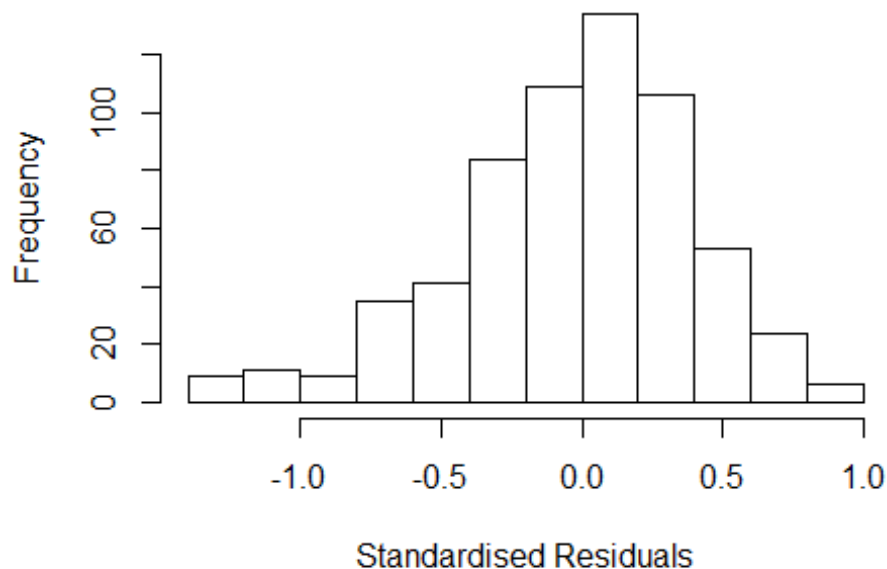
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3898 -0.2507 0.0271 0.2600 0.9940
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25219 0.06258 20.01 <2e-16 ***
## FirstAuthorFemale1 -0.01616 0.03506 -0.46 0.645
## LastAuthorFemale1 -0.00216 0.04314 -0.05 0.960
## Year1997 0.01855 0.09070 0.20 0.838
## Year1998 -0.01726 0.09079 -0.19 0.849
## Year1999 -0.16187 0.09707 -1.67 0.096 .
## Year2000 -0.02865 0.11457 -0.25 0.803
## Year2001 0.03612 0.08567 0.42 0.673
## Year2002 -0.18673 0.08845 -2.11 0.035 *
## Year2003 -0.16335 0.11138 -1.47 0.143
## Year2004 -0.14071 0.08730 -1.61 0.108
## Year2005 -0.09050 0.08725 -1.04 0.300
```

```

## Year2006      -0.05275    0.08414   -0.63    0.531
## Year2007      0.15598    0.09555    1.63    0.103
## Year2008     -0.03936    0.08803   -0.45    0.655
## Year2009      0.06167    0.08787    0.70    0.483
## Year2010      0.04368    0.09045    0.48    0.629
## Year2011     -0.00400    0.09061   -0.04    0.965
## Year2012      0.08582    0.09877    0.87    0.385
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0513, Adjusted R-squared:  0.023
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 579 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.180  0.869  0.952  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.61e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.145 1      1.070
## Year              1.145 16      1.004

```

## Residuals from first author

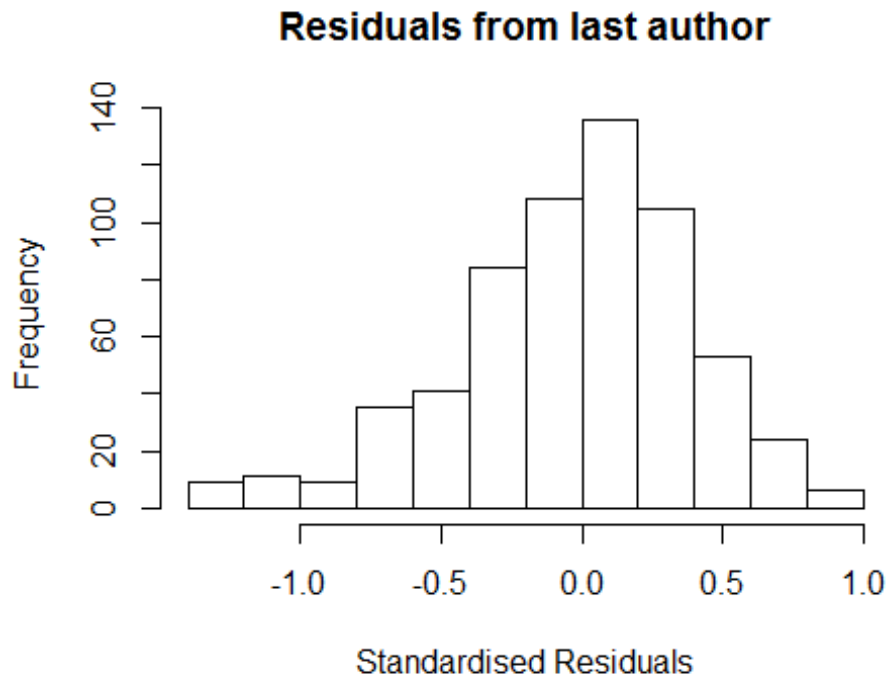


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3919 -0.2505 0.0272 0.2601 0.9943
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25199 0.06234 20.08 <2e-16 ***
## FirstAuthorFemale1 -0.01636 0.03505 -0.47 0.641
## Year1997 0.01851 0.09054 0.20 0.838
## Year1998 -0.01723 0.09077 -0.19 0.849
## Year1999 -0.16212 0.09652 -1.68 0.094 .
## Year2000 -0.02850 0.11447 -0.25 0.804
## Year2001 0.03581 0.08528 0.42 0.675
## Year2002 -0.18682 0.08830 -2.12 0.035 *
## Year2003 -0.16337 0.11129 -1.47 0.143
## Year2004 -0.14090 0.08743 -1.61 0.108
## Year2005 -0.09058 0.08742 -1.04 0.301
## Year2006 -0.05263 0.08401 -0.63 0.531
```

```

## Year2007          0.15632      0.09529      1.64      0.101
## Year2008          -0.03943      0.08802     -0.45      0.654
## Year2009           0.06173      0.08787      0.70      0.483
## Year2010           0.04365      0.09044      0.48      0.630
## Year2011          -0.00421      0.09000     -0.05      0.963
## Year2012           0.08570      0.09846      0.87      0.384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.0514, Adjusted R-squared:  0.0246
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 576 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.177  0.868  0.952  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.61e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.11 1      1.053
## Year              1.11 16      1.003

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3932 -0.2590  0.0244  0.2607  1.0006
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.24719    0.06142   20.31  <2e-16 ***
## LastAuthorFemale1 -0.00378    0.04325   -0.09    0.930
## Year1997         0.01762    0.09133    0.19    0.847
## Year1998        -0.01692    0.09090   -0.19    0.852
## Year1999        -0.16032    0.09685   -1.66    0.098 .
## Year2000        -0.03124    0.11363   -0.27    0.783
## Year2001         0.03474    0.08564    0.41    0.685
## Year2002        -0.18879    0.08845   -2.13    0.033 *
## Year2003        -0.16642    0.11179   -1.49    0.137
## Year2004        -0.13902    0.08755   -1.59    0.113
## Year2005        -0.09160    0.08718   -1.05    0.294
## Year2006        -0.05333    0.08432   -0.63    0.527
```

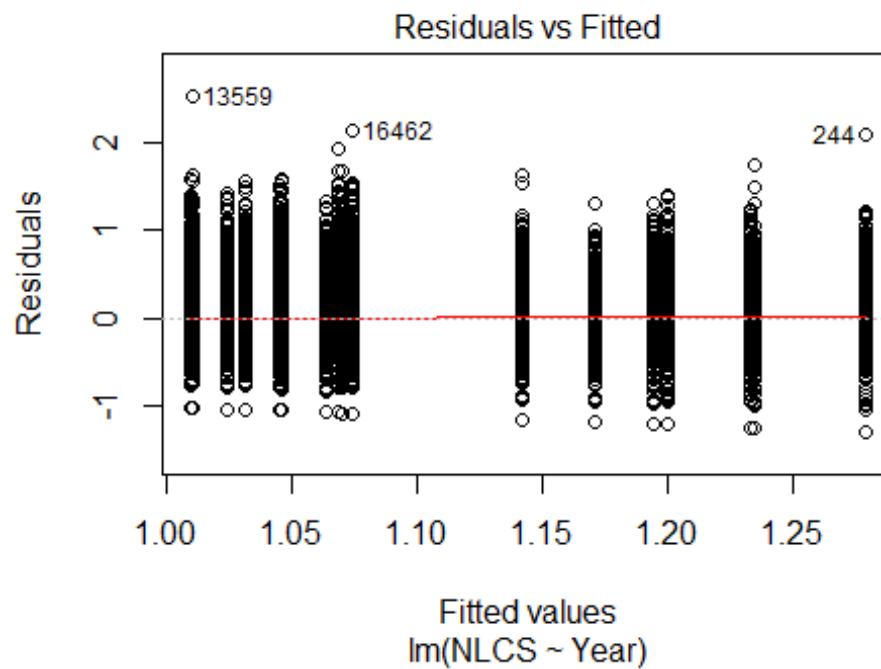


```

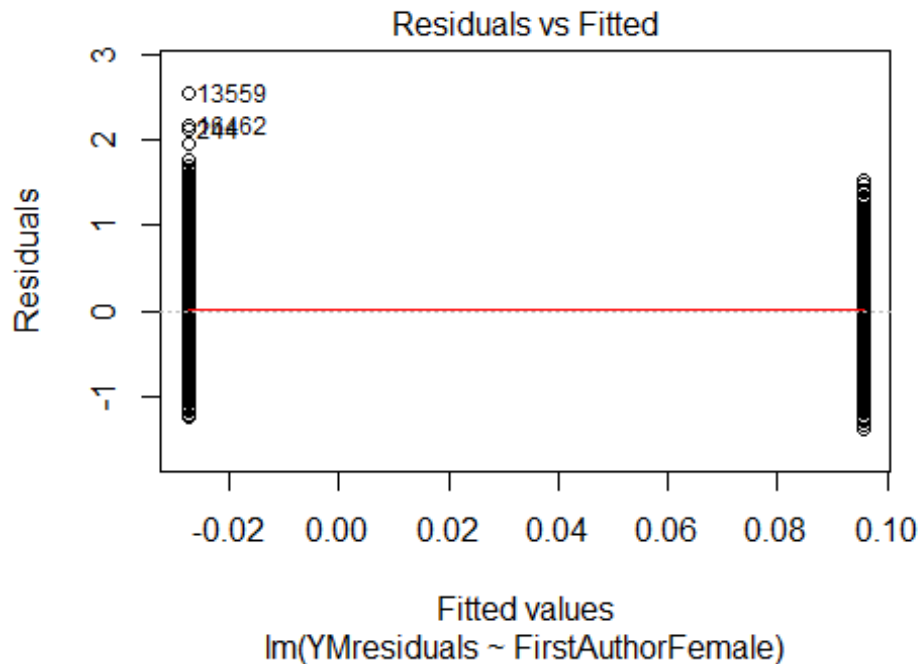
## Year2007      0.14981    0.09397    1.59    0.111
## Year2008     -0.04023    0.08835   -0.46    0.649
## Year2009      0.06105    0.08813    0.69    0.489
## Year2010      0.04113    0.09084    0.45    0.651
## Year2011     -0.00516    0.09073   -0.06    0.955
## Year2012      0.08423    0.09925    0.85    0.396
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.0507, Adjusted R-squared:  0.0239
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 44 weights are ~= 1. The remaining 577 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.180  0.871  0.953  0.898  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.61e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 621"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2705"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  803 1040 1000  973 1205 1036  982  800  916  889 1019 1054 1028 1020 1029
## 2011 2012
##  991  950
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  336  388  340  292  475  417  557  448  594  580  690  692  674  709  726
## 2011 2012

```

```
## 722 684
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 299 334 301 263 425 368 466 370 510 463 582 597 568 599 628
## 2011 2012
## 615 582
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 63, df = 16, p-value = 2e-07
```

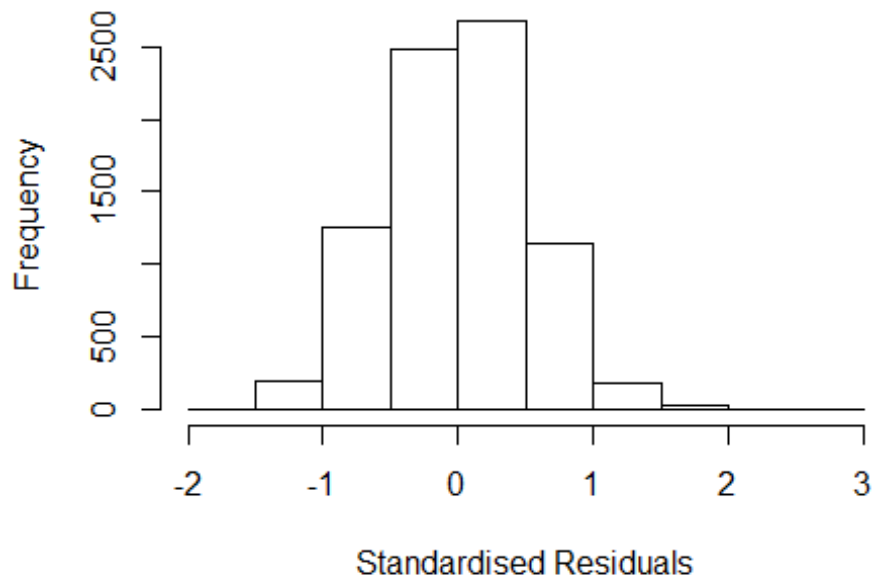


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 5.46660431048482"
## [1] "Male first author team size 2018 geometric mean: 4.78193221417128"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 6.01725387283824"
## [1] "Male last author team size 2018 geometric mean: 4.78265491803296"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.050 1          1.025
## LastAuthorFemale  1.042 1          1.021
## UniqueAuthors    1.069 4          1.008
## Year              1.084 16         1.003
```

## Residuals from first and last author and team size



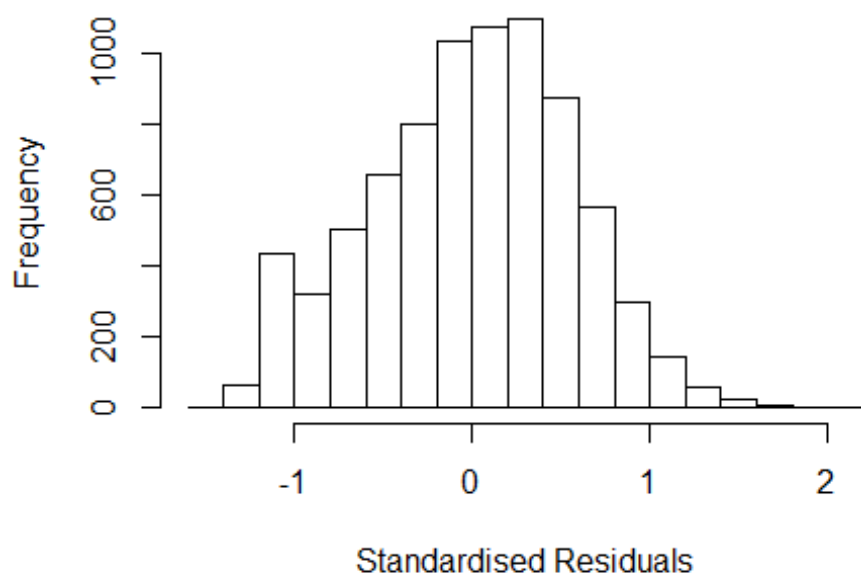
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 244 0029928495 3.368 1996      2705      2      2.58
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57148 -0.36604  0.00802  0.36055  2.57952
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7885    0.0396   19.91 < 2e-16 ***
## FirstAuthorFemale1 0.1139    0.0146    7.81 6.2e-15 ***
## LastAuthorFemale1 0.0717    0.0191    3.74 0.00018 ***
## UniqueAuthors2    0.2850    0.0321    8.88 < 2e-16 ***
## UniqueAuthors3    0.3137    0.0294   10.67 < 2e-16 ***
## UniqueAuthors4    0.4630    0.0289   16.05 < 2e-16 ***
## UniqueAuthors5    0.7113    0.0266   26.69 < 2e-16 ***
## Year1997        -0.0493    0.0428   -1.15 0.24954
## Year1998        -0.0555    0.0452   -1.23 0.21888
## Year1999        -0.0425    0.0454   -0.94 0.34969
```

```

## Year2000          -0.0490      0.0402   -1.22   0.22200
## Year2001          -0.1200      0.0400   -3.00   0.00270 **
## Year2002          -0.1627      0.0390   -4.17   3.1e-05 ***
## Year2003          -0.2524      0.0406   -6.22   5.1e-10 ***
## Year2004          -0.2833      0.0383   -7.40   1.5e-13 ***
## Year2005          -0.2652      0.0394   -6.73   1.8e-11 ***
## Year2006          -0.2723      0.0394   -6.91   5.1e-12 ***
## Year2007          -0.2717      0.0383   -7.10   1.4e-12 ***
## Year2008          -0.2962      0.0393   -7.54   5.2e-14 ***
## Year2009          -0.3109      0.0376   -8.26   < 2e-16 ***
## Year2010          -0.2815      0.0399   -7.05   1.9e-12 ***
## Year2011          -0.2747      0.0390   -7.05   2.0e-12 ***
## Year2012          -0.2676      0.0401   -6.67   2.8e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.528
## Multiple R-squared:  0.18, Adjusted R-squared:  0.177
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 107 is an outlier with |weight| = 0 ( < 1.3e-05);
## 657 weights are ~= 1. The remaining 7312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0322 0.8710 0.9490 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.25e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1          1.011
## LastAuthorFemale 1.025 1          1.012
## Year 1.022 16          1.001

```

## Residuals from first and last author



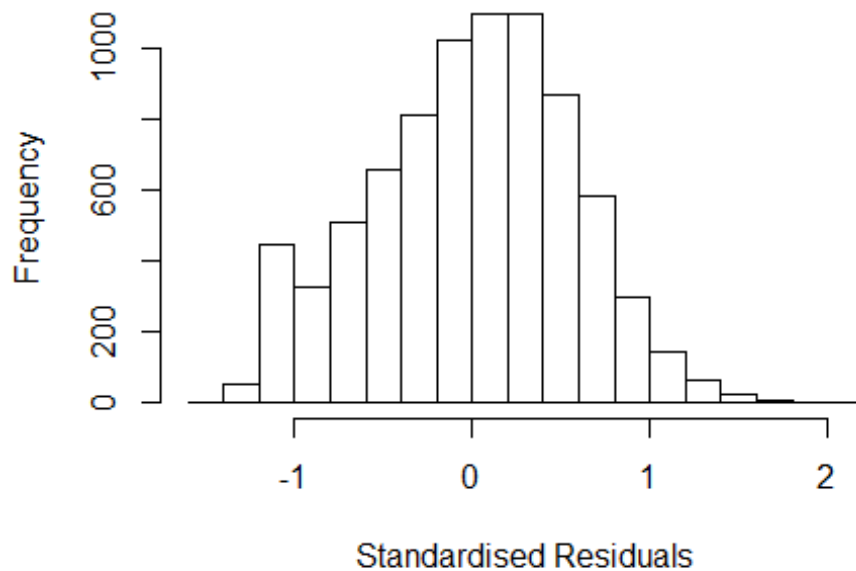
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4549 -0.3965 0.0281 0.3944 2.1765
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2694 0.0331 38.34 < 2e-16 ***
## FirstAuthorFemale1 0.1382 0.0155 8.94 < 2e-16 ***
## LastAuthorFemale1 0.0473 0.0208 2.27 0.0233 *
## Year1997 -0.0647 0.0450 -1.44 0.1509
## Year1998 -0.0870 0.0481 -1.81 0.0706 .
## Year1999 -0.0675 0.0482 -1.40 0.1616
## Year2000 -0.0828 0.0423 -1.96 0.0504 .
## Year2001 -0.1295 0.0418 -3.10 0.0019 **
## Year2002 -0.1539 0.0414 -3.72 0.0002 ***
## Year2003 -0.2364 0.0441 -5.36 8.4e-08 ***
## Year2004 -0.2967 0.0425 -6.99 3.1e-12 ***
## Year2005 -0.2832 0.0430 -6.59 4.7e-11 ***
```

```

## Year2006          -0.2633      0.0426   -6.17  6.9e-10 ***
## Year2007          -0.2569      0.0414   -6.21  5.6e-10 ***
## Year2008          -0.2885      0.0418   -6.91  5.3e-12 ***
## Year2009          -0.2833      0.0402   -7.05  2.0e-12 ***
## Year2010          -0.2349      0.0430   -5.46  5.0e-08 ***
## Year2011          -0.2298      0.0416   -5.52  3.6e-08 ***
## Year2012          -0.2298      0.0433   -5.31  1.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0313, Adjusted R-squared:  0.0291
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 651 weights are ~= 1. The remaining 7319 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.122  0.867  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##      trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.01 1      1.005
## Year      1.01 16      1.000

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4153 -0.3955 0.0292 0.3967 2.1714
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2718 0.0331 38.42 < 2e-16 ***
## FirstAuthorFemale1 0.1435 0.0154 9.34 < 2e-16 ***
## Year1997 -0.0635 0.0450 -1.41 0.15863
## Year1998 -0.0850 0.0481 -1.77 0.07692 .
## Year1999 -0.0652 0.0481 -1.35 0.17594
## Year2000 -0.0824 0.0423 -1.95 0.05139 .
## Year2001 -0.1279 0.0418 -3.06 0.00219 **
## Year2002 -0.1524 0.0413 -3.69 0.00023 ***
## Year2003 -0.2369 0.0441 -5.37 7.9e-08 ***
## Year2004 -0.2940 0.0424 -6.94 4.2e-12 ***
## Year2005 -0.2819 0.0429 -6.56 5.6e-11 ***
## Year2006 -0.2615 0.0426 -6.13 9.0e-10 ***
```

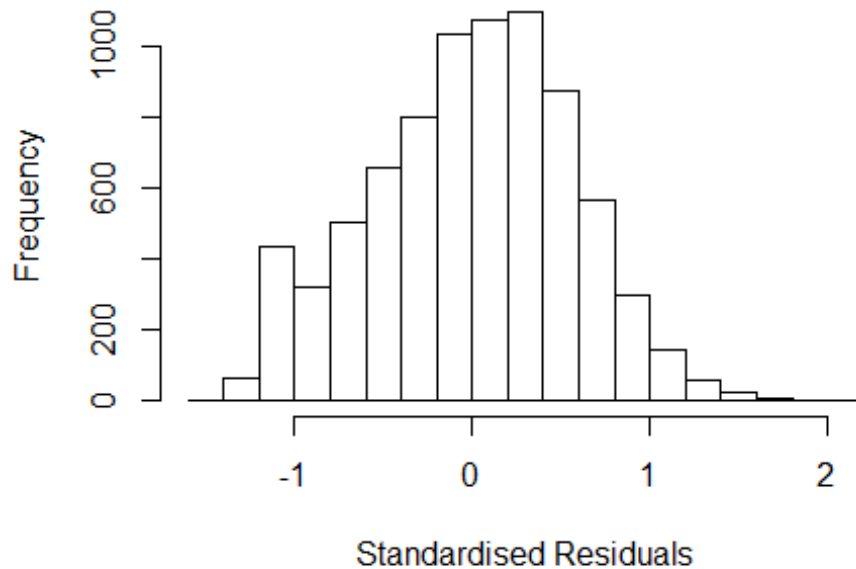


```

## Year2007          -0.2547      0.0413   -6.16  7.6e-10 ***
## Year2008          -0.2873      0.0417   -6.88  6.2e-12 ***
## Year2009          -0.2813      0.0402   -7.00  2.7e-12 ***
## Year2010          -0.2322      0.0430   -5.39  7.1e-08 ***
## Year2011          -0.2280      0.0416   -5.48  4.4e-08 ***
## Year2012          -0.2286      0.0433   -5.28  1.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0306, Adjusted R-squared:  0.0285
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 648 weights are ~= 1. The remaining 7322 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.124  0.867  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.013 1      1.006
## Year              1.013 16      1.000

```

## Residuals from last author



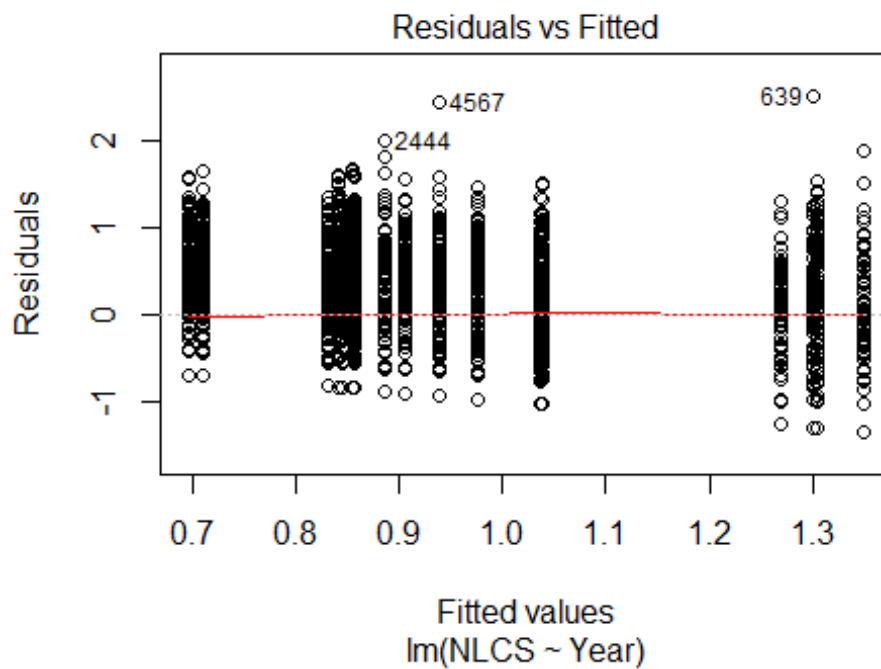
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3690 -0.4033  0.0303  0.3950  2.1439
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2939     0.0330   39.22 < 2e-16 ***
## LastAuthorFemale1  0.0751     0.0207    3.63  0.00028 ***
## Year1997         -0.0645     0.0451   -1.43  0.15283
## Year1998         -0.0843     0.0481   -1.75  0.08010 .
## Year1999         -0.0618     0.0481   -1.29  0.19874
## Year2000         -0.0835     0.0425   -1.97  0.04939 *
## Year2001         -0.1265     0.0418   -3.02  0.00250 **
## Year2002         -0.1516     0.0414   -3.66  0.00025 ***
## Year2003         -0.2324     0.0443   -5.24  1.6e-07 ***
## Year2004         -0.2960     0.0427   -6.94  4.3e-12 ***
## Year2005         -0.2821     0.0431   -6.54  6.4e-11 ***
## Year2006         -0.2670     0.0427   -6.26  4.2e-10 ***
```

```

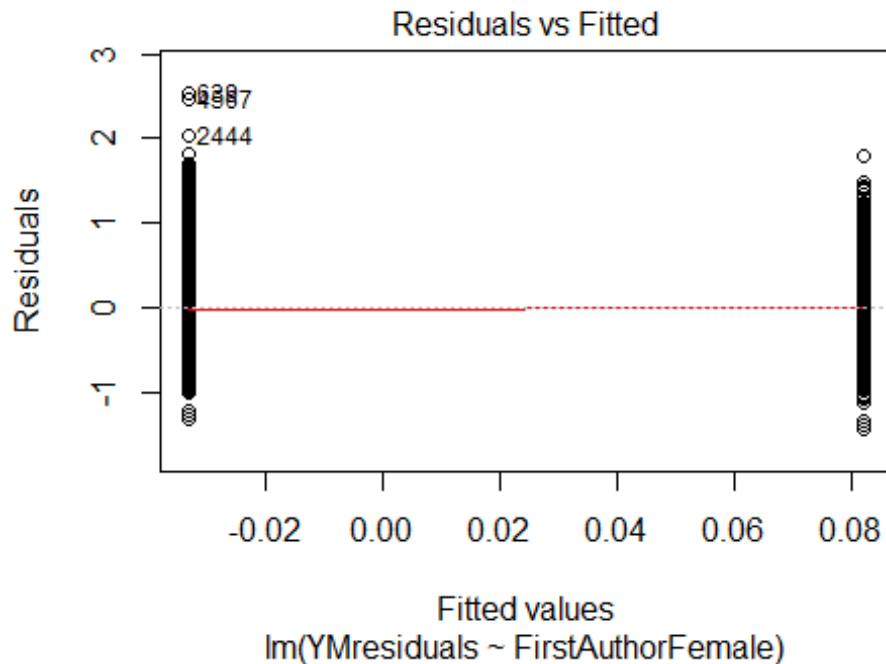
## Year2007          -0.2528      0.0416   -6.08  1.3e-09 ***
## Year2008          -0.2792      0.0419   -6.67  2.7e-11 ***
## Year2009          -0.2767      0.0403   -6.87  7.0e-12 ***
## Year2010          -0.2268      0.0432   -5.25  1.5e-07 ***
## Year2011          -0.2240      0.0417   -5.37  8.0e-08 ***
## Year2012          -0.2183      0.0433   -5.05  4.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.58
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.0198
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 647 weights are ~= 1. The remaining 7323 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.142  0.863  0.949  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7970"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2706"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 280 272 271 234 327 372 338 273 407 449 390 384 370 365 395
## 2011 2012
## 384 330
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 80 83 79 108 145 150 164 163 265 296 228 244 222 170 279
## 2011 2012

```

```
## 279 240
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 75 76 68 97 137 126 135 123 216 223 169 205 163 138 230
## 2011 2012
## 230 205
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```

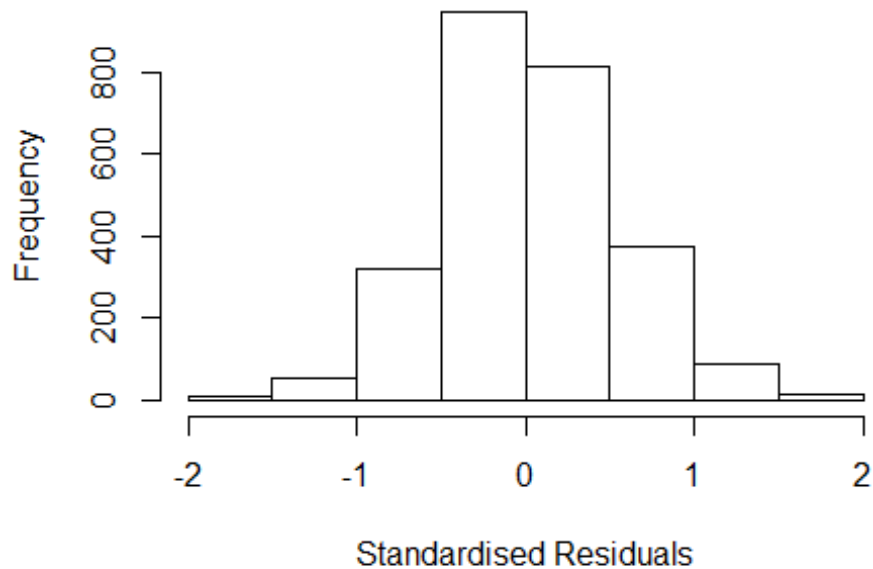


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.02, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 3.78274276101071"
## [1] "Male first author team size 2018 geometric mean: 3.76380510598294"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.90234373252955"
## [1] "Male last author team size 2018 geometric mean: 3.72757592253137"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.298 1          1.140
## LastAuthorFemale  1.251 1          1.119
## UniqueAuthors    1.367 4          1.040
## Year              1.416 16         1.011
```

## Residuals from first and last author and team size



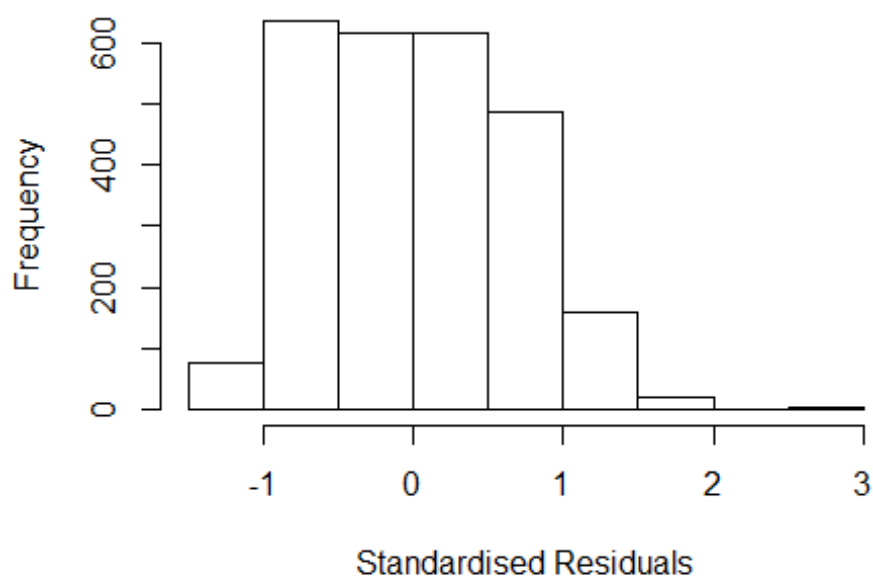
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.76826 -0.34195 -0.00633 0.35860 1.96923
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8342 0.0643 12.97 < 2e-16 ***
## FirstAuthorFemale1 0.0864 0.0263 3.29 0.001 **
## LastAuthorFemale1 0.0299 0.0302 0.99 0.322
## UniqueAuthors2 0.1989 0.0308 6.46 1.2e-10 ***
## UniqueAuthors3 0.5342 0.0421 12.69 < 2e-16 ***
## UniqueAuthors4 0.8366 0.0391 21.41 < 2e-16 ***
## UniqueAuthors5 1.0869 0.0282 38.54 < 2e-16 ***
## Year1997 -0.2168 0.0911 -2.38 0.017 *
## Year1998 -0.0634 0.1056 -0.60 0.548
## Year1999 -0.1529 0.0871 -1.76 0.079 .
```

```

## Year2000      -0.4067      0.0796      -5.11      3.5e-07 ***
## Year2001      -0.4518      0.0766      -5.90      4.2e-09 ***
## Year2002      -0.4995      0.0809      -6.17      7.8e-10 ***
## Year2003      -0.5322      0.0803      -6.63      4.2e-11 ***
## Year2004      -0.5600      0.0691      -8.11      8.0e-16 ***
## Year2005      -0.5666      0.0692      -8.18      4.3e-16 ***
## Year2006      -0.5428      0.0729      -7.44      1.3e-13 ***
## Year2007      -0.4194      0.0732      -5.73      1.1e-08 ***
## Year2008      -0.5563      0.0828      -6.72      2.3e-11 ***
## Year2009      -0.4407      0.0844      -5.22      1.9e-07 ***
## Year2010      -0.5597      0.0739      -7.57      5.0e-14 ***
## Year2011      -0.5196      0.0727      -7.15      1.1e-12 ***
## Year2012      -0.5734      0.0735      -7.81      8.5e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.455, Adjusted R-squared:  0.45
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 222 weights are ~= 1. The remaining 2394 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0609 0.8620 0.9430 0.8920 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.82e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.197 1      1.094
## LastAuthorFemale  1.181 1      1.087
## Year              1.056 16      1.002

```

## Residuals from first and last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 639   0031746535 3.800 1998    2700      3    2.573
## 4567 48249085691 3.364 2008    2700      3    2.545
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4457 -0.5648 -0.0125  0.5150  2.5730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3185    0.0671   19.66 < 2e-16 ***
## FirstAuthorFemale1  0.1640    0.0329    4.98 6.8e-07 ***
## LastAuthorFemale1 -0.0368    0.0375   -0.98 0.32695
## Year1997          -0.0332    0.0947   -0.35 0.72559
## Year1998          -0.0547    0.1154   -0.47 0.63575
## Year1999          -0.0519    0.1096   -0.47 0.63587
## Year2000          -0.4818    0.0946   -5.10 3.7e-07 ***
## Year2001          -0.4978    0.0945   -5.26 1.5e-07 ***
## Year2002          -0.3125    0.0906   -3.45 0.00057 ***
## Year2003          -0.5410    0.1013   -5.34 1.0e-07 ***
## Year2004          -0.7432    0.0810   -9.17 < 2e-16 ***
```

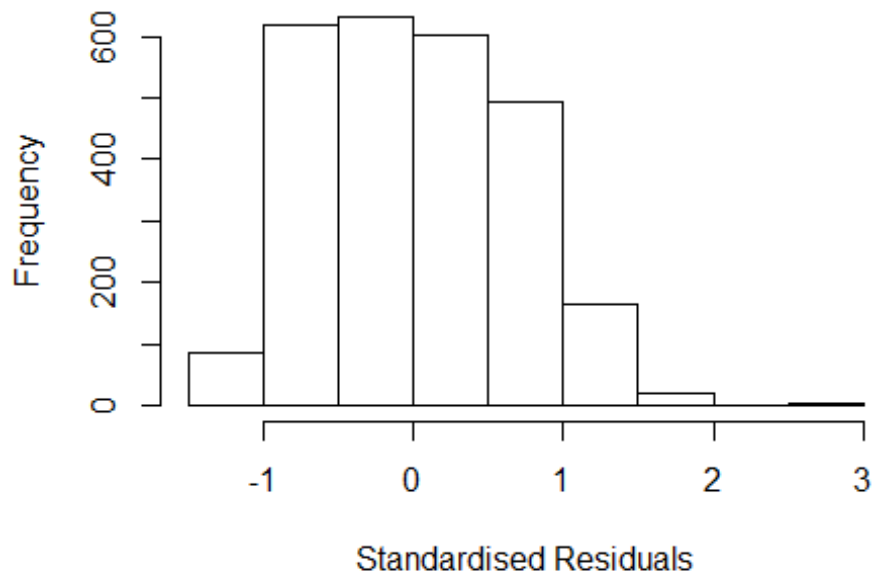


```

## Year2005          -0.7460      0.0795    -9.39 < 2e-16 ***
## Year2006          -0.5857      0.0835    -7.01 3.0e-12 ***
## Year2007          -0.3527      0.0805    -4.38 1.2e-05 ***
## Year2008          -0.4626      0.0887    -5.22 2.0e-07 ***
## Year2009          -0.2961      0.0912    -3.25 0.00118 **
## Year2010          -0.4965      0.0817    -6.08 1.4e-09 ***
## Year2011          -0.5082      0.0822    -6.19 7.1e-10 ***
## Year2012          -0.5457      0.0871    -6.27 4.3e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0896, Adjusted R-squared:  0.0833
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~ = 1. The remaining 2444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.164  0.883   0.942   0.921   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0          1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1          1.019
## Year              1.038 16          1.001

```

## Residuals from first author



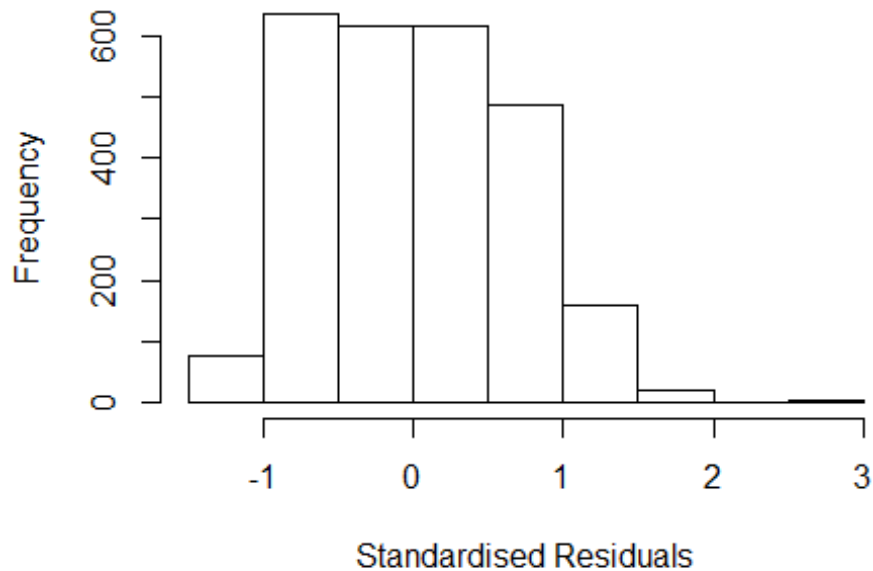
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 639   0031746535 3.800 1998    2700      3    2.573
## 4567 48249085691 3.364 2008    2700      3    2.545
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.47086 -0.56525 -0.00864  0.51424  2.53943
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3174     0.0671   19.62 < 2e-16 ***
## FirstAuthorFemale1 0.1535     0.0308    4.98 6.6e-07 ***
## Year1997        -0.0345     0.0947   -0.36 0.71540
## Year1998        -0.0568     0.1160   -0.49 0.62437
## Year1999        -0.0540     0.1098   -0.49 0.62260
## Year2000        -0.4849     0.0950   -5.11 3.5e-07 ***
## Year2001        -0.5010     0.0948   -5.29 1.3e-07 ***
## Year2002        -0.3155     0.0908   -3.47 0.00052 ***
## Year2003        -0.5456     0.1013   -5.39 7.8e-08 ***
## Year2004        -0.7469     0.0812   -9.20 < 2e-16 ***
## Year2005        -0.7487     0.0795   -9.41 < 2e-16 ***
```

```

## Year2006          -0.5912      0.0836   -7.07   2.0e-12 ***
## Year2007          -0.3553      0.0806   -4.41   1.1e-05 ***
## Year2008          -0.4635      0.0887   -5.23   1.9e-07 ***
## Year2009          -0.2987      0.0913   -3.27   0.00108 **
## Year2010          -0.5015      0.0816   -6.14   9.4e-10 ***
## Year2011          -0.5129      0.0822   -6.24   5.0e-10 ***
## Year2012          -0.5501      0.0870   -6.32   3.0e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0892, Adjusted R-squared:  0.0832
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 167 weights are ~= 1. The remaining 2449 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
## 0.177 0.884 0.943 0.921 0.983 0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.010
## Year      1.021 16      1.001

```

## Residuals from last author



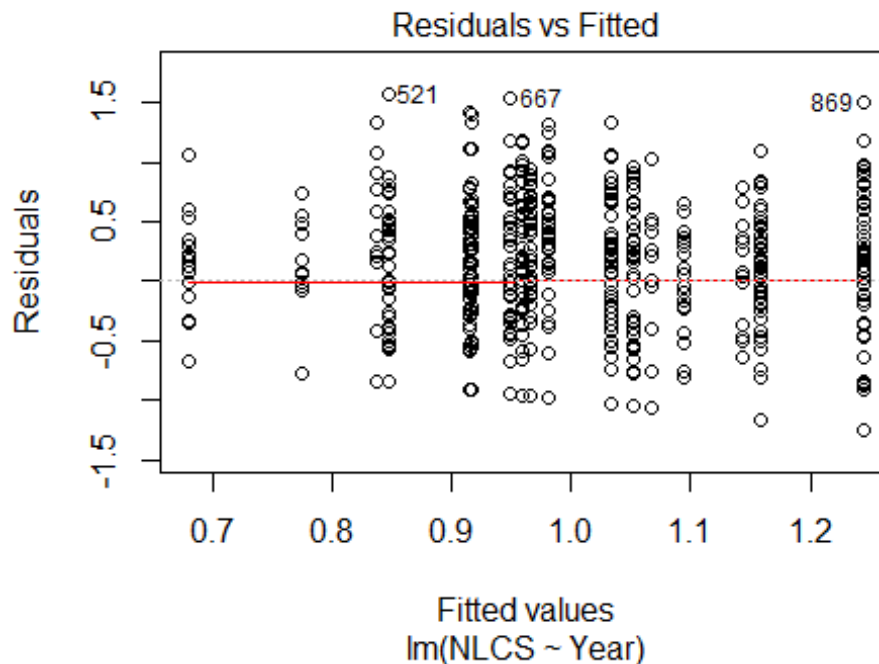
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 639   0031746535 3.800 1998    2700      3    2.573
## 4567 48249085691 3.364 2008    2700      3    2.545
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.36484 -0.57020 -0.00938  0.52170  2.48813
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3416     0.0675   19.88 < 2e-16 ***
## LastAuthorFemale1  0.0233     0.0342    0.68  0.4963
## Year1997        -0.0321     0.0947   -0.34  0.7349
## Year1998        -0.0530     0.1156   -0.46  0.6470
## Year1999        -0.0406     0.1099   -0.37  0.7121
## Year2000        -0.4567     0.0959  -4.76 2.0e-06 ***
## Year2001        -0.4887     0.0959  -5.10 3.7e-07 ***
## Year2002        -0.2880     0.0911  -3.16  0.0016 **
## Year2003        -0.5298     0.1014  -5.23 1.9e-07 ***
## Year2004        -0.7394     0.0815  -9.07 < 2e-16 ***
## Year2005        -0.7364     0.0800  -9.21 < 2e-16 ***
```

```

## Year2006          -0.5725      0.0834    -6.87  8.1e-12 ***
## Year2007          -0.3361      0.0811    -4.15  3.5e-05 ***
## Year2008          -0.4388      0.0887    -4.94  8.1e-07 ***
## Year2009          -0.2803      0.0920    -3.05  0.0023 **
## Year2010          -0.4844      0.0824    -5.88  4.7e-09 ***
## Year2011          -0.4963      0.0827    -6.00  2.3e-09 ***
## Year2012          -0.5359      0.0871    -6.15  8.7e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.712
## Multiple R-squared:  0.0798, Adjusted R-squared:  0.0738
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 199 weights are ~= 1. The remaining 2417 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.197  0.876  0.940  0.919  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##      trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2616"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2707"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   85   84   71   86   76   74   60   58   55   59   75   78   84   70   65
## 2011 2012
##   68   61
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   14   17   13   22   22   38   41   40   43   49   55   67   55   51

```

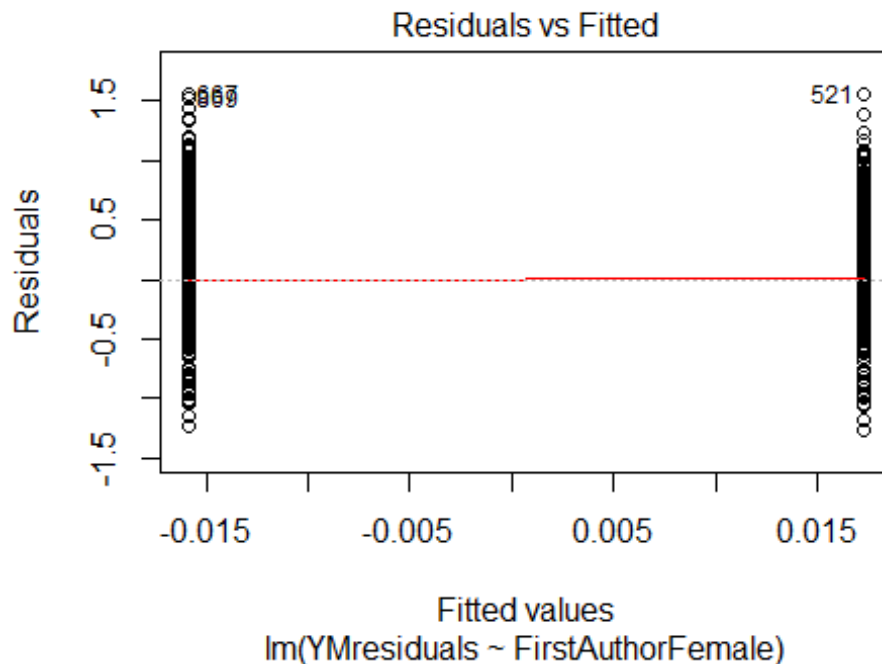
```
## 2011 2012
## 57 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 18 13 16 12 20 21 36 33 38 37 46 49 62 52 43
## 2011 2012
## 51 49
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.35, df = 1, p-value = 0.6
## [1] "Female first author team size 2018 geometric mean: 3.48804982369966"
## [1] "Male first author team size 2018 geometric mean: 2.48569135760778"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

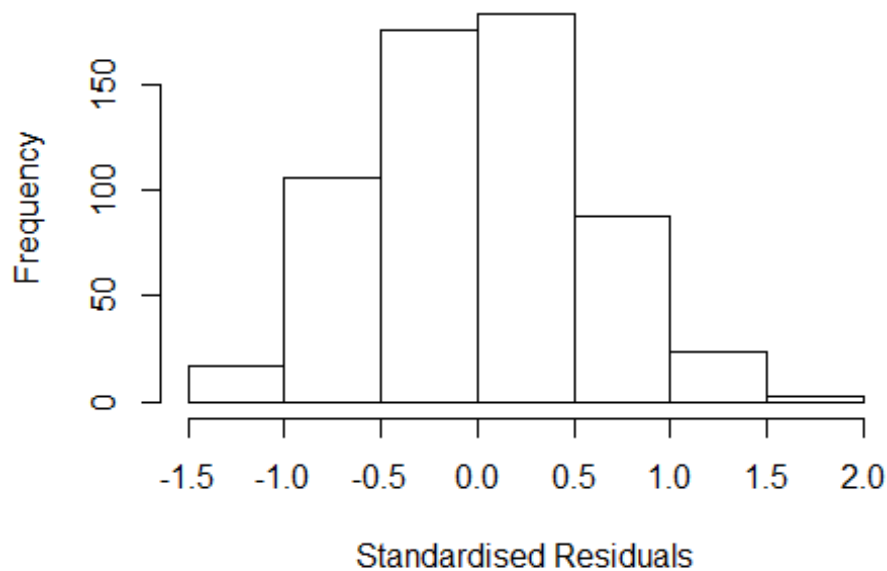
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 220, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.37503562634741"
## [1] "Male last author team size 2018 geometric mean: 2.54885852519547"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.771 1      1.331
## LastAuthorFemale  1.751 1      1.323
## UniqueAuthors    1.944 4      1.087
## Year              2.149 16     1.024
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.39879 -0.38456 0.00167 0.38434 1.79948
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.51593 0.17459 2.96 0.0033 **
## FirstAuthorFemale1 -0.00285 0.06230 -0.05 0.9635
## LastAuthorFemale1 -0.05052 0.06422 -0.79 0.4318
## UniqueAuthors2 0.17637 0.07510 2.35 0.0192 *
## UniqueAuthors3 0.48894 0.07718 6.34 4.8e-10 ***
## UniqueAuthors4 0.59952 0.07990 7.50 2.4e-13 ***
## UniqueAuthors5 0.63805 0.07486 8.52 < 2e-16 ***
## Year1997 0.26822 0.24416 1.10 0.2724
## Year1998 0.45089 0.19788 2.28 0.0231 *
## Year1999 0.12499 0.21139 0.59 0.5546
```

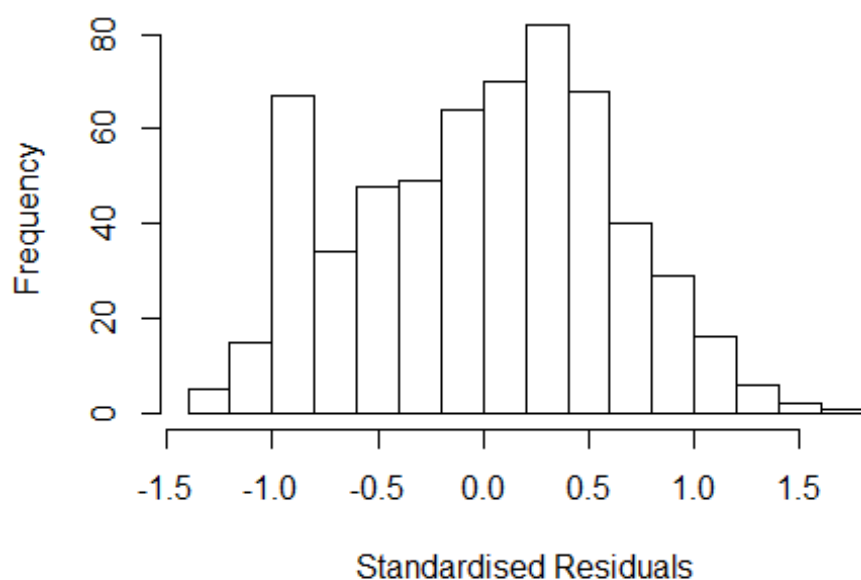


```

## Year2000          0.50303      0.19398      2.59      0.0098 **
## Year2001          0.06174      0.19944      0.31      0.7570
## Year2002          0.15497      0.19252      0.80      0.4212
## Year2003          0.26137      0.20268      1.29      0.1977
## Year2004          0.24671      0.18742      1.32      0.1886
## Year2005          0.25328      0.18777      1.35      0.1779
## Year2006          0.48943      0.19752      2.48      0.0135 *
## Year2007          0.46284      0.20325      2.28      0.0231 *
## Year2008          0.24965      0.18750      1.33      0.1836
## Year2009          0.21002      0.19114      1.10      0.2723
## Year2010          0.21585      0.19968      1.08      0.2802
## Year2011          0.24481      0.19551      1.25      0.2110
## Year2012          0.01754      0.18689      0.09      0.9253
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.583
## Multiple R-squared:  0.198, Adjusted R-squared:  0.167
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 555 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.320  0.870  0.951  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.68e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.645 1      1.283
## LastAuthorFemale  1.674 1      1.294
## Year              1.220 16      1.006

```

## Residuals from first and last author



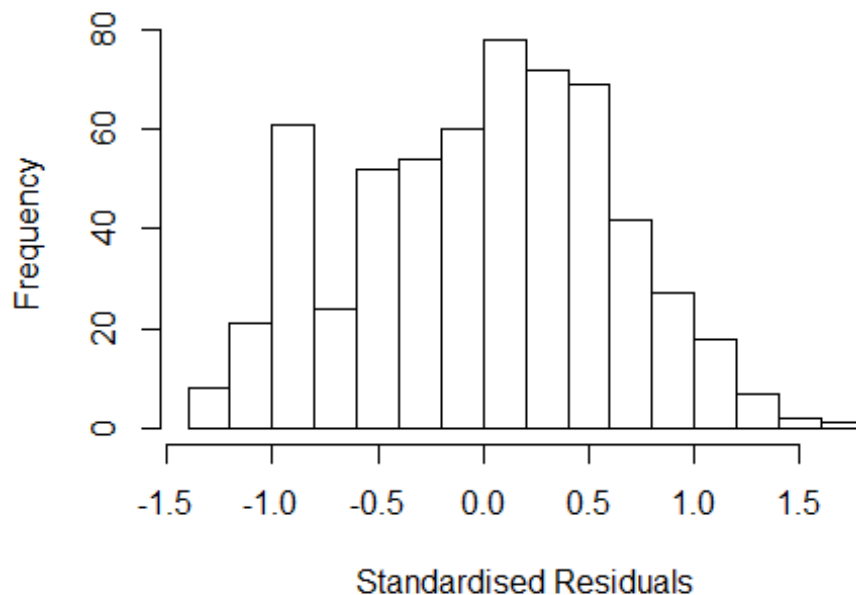
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2825 -0.4953 0.0437 0.4279 1.6084
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8193 0.2233 3.67 0.00027 ***
## FirstAuthorFemale1 0.0801 0.0664 1.21 0.22835
## LastAuthorFemale1 -0.1073 0.0694 -1.55 0.12247
## Year1997 0.2632 0.2907 0.91 0.36573
## Year1998 0.3490 0.2486 1.40 0.16098
## Year1999 0.0342 0.2595 0.13 0.89507
## Year2000 0.2666 0.2405 1.11 0.26809
## Year2001 -0.1504 0.2494 -0.60 0.54656
## Year2002 0.0165 0.2435 0.07 0.94608
## Year2003 0.1313 0.2504 0.52 0.60006
## Year2004 0.0776 0.2470 0.31 0.75362
## Year2005 0.0381 0.2382 0.16 0.87303
```

```

## Year2006          0.3831      0.2445      1.57  0.11768
## Year2007          0.4485      0.2481      1.81  0.07121 .
## Year2008          0.2027      0.2376      0.85  0.39385
## Year2009          0.1552      0.2482      0.63  0.53206
## Year2010          0.1324      0.2512      0.53  0.59831
## Year2011          0.1997      0.2429      0.82  0.41119
## Year2012          0.0859      0.2421      0.35  0.72277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.0527, Adjusted R-squared:  0.0232
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 553 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.533  0.868  0.949  0.920  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.68e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1          1.043
## Year              1.087 16          1.003

```

## Residuals from first author



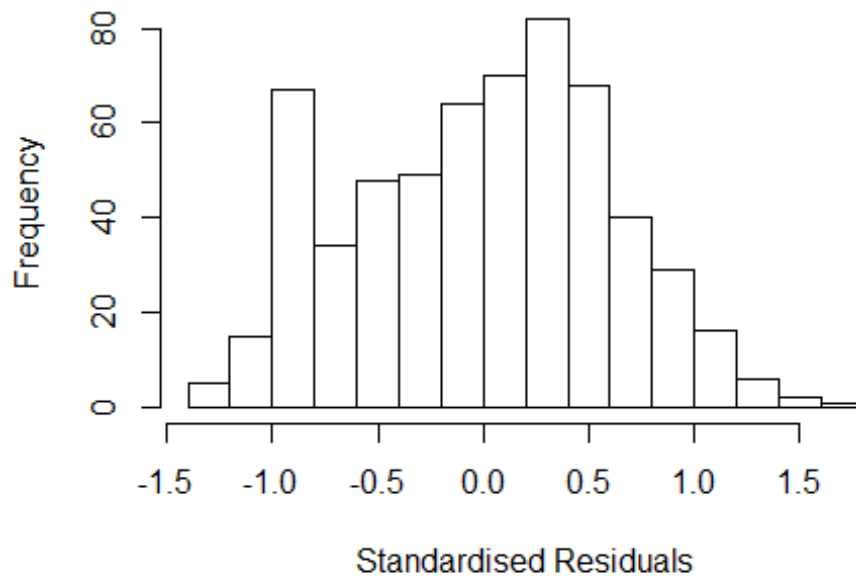
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2750 -0.4876 0.0475 0.4285 1.6070
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.81320 0.22293 3.65 0.00029 ***
## FirstAuthorFemale1 0.02461 0.05436 0.45 0.65088
## Year1997 0.25523 0.28851 0.88 0.37672
## Year1998 0.35644 0.24914 1.43 0.15306
## Year1999 0.02617 0.25923 0.10 0.91961
## Year2000 0.26561 0.24006 1.11 0.26899
## Year2001 -0.16407 0.25001 -0.66 0.51191
## Year2002 -0.00917 0.24240 -0.04 0.96985
## Year2003 0.12086 0.24775 0.49 0.62586
## Year2004 0.06777 0.24850 0.27 0.78517
## Year2005 0.01637 0.23692 0.07 0.94493
## Year2006 0.36521 0.24192 1.51 0.13169
```

```

## Year2007          0.43715      0.24777      1.76  0.07821 .
## Year2008          0.19939      0.23729      0.84  0.40110
## Year2009          0.13492      0.24795      0.54  0.58654
## Year2010          0.12209      0.25014      0.49  0.62566
## Year2011          0.19893      0.24215      0.82  0.41168
## Year2012          0.08938      0.24246      0.37  0.71252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.662
## Multiple R-squared:  0.0487, Adjusted R-squared:  0.0207
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 548 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.535  0.867   0.950   0.920   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.68e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.116 1      1.056
## Year              1.116 16      1.003

```

## Residuals from last author



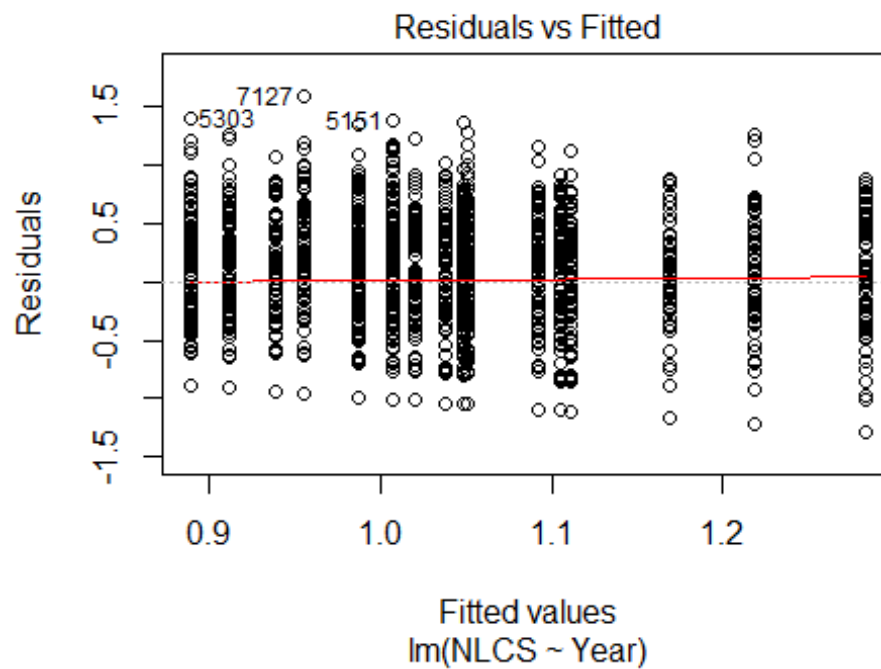
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2901 -0.4976 0.0372 0.4346 1.6267
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8334 0.2236 3.73 0.00021 ***
## LastAuthorFemale1 -0.0633 0.0565 -1.12 0.26316
## Year1997 0.2602 0.2916 0.89 0.37270
## Year1998 0.3485 0.2508 1.39 0.16520
## Year1999 0.0422 0.2577 0.16 0.86989
## Year2000 0.2757 0.2418 1.14 0.25464
## Year2001 -0.1451 0.2503 -0.58 0.56227
## Year2002 0.0202 0.2439 0.08 0.93387
## Year2003 0.1302 0.2499 0.52 0.60256
## Year2004 0.0813 0.2474 0.33 0.74254
## Year2005 0.0422 0.2395 0.18 0.86007
## Year2006 0.3879 0.2449 1.58 0.11380
```

```

## Year2007          0.4568      0.2492      1.83  0.06738 .
## Year2008          0.2143      0.2387      0.90  0.36968
## Year2009          0.1557      0.2495      0.62  0.53282
## Year2010          0.1414      0.2522      0.56  0.57522
## Year2011          0.2137      0.2438      0.88  0.38122
## Year2012          0.1037      0.2435      0.43  0.67030
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.0502, Adjusted R-squared:  0.0222
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 542 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.523  0.866  0.945  0.918  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.68e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 596"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2708"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 457 469 400 384 396 390 342 341 342 321 389 368 350 390 309
## 2011 2012
## 316 266
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 110 92 83 82 65 108 114 140 138 146 168 152 121 105
## 2011 2012

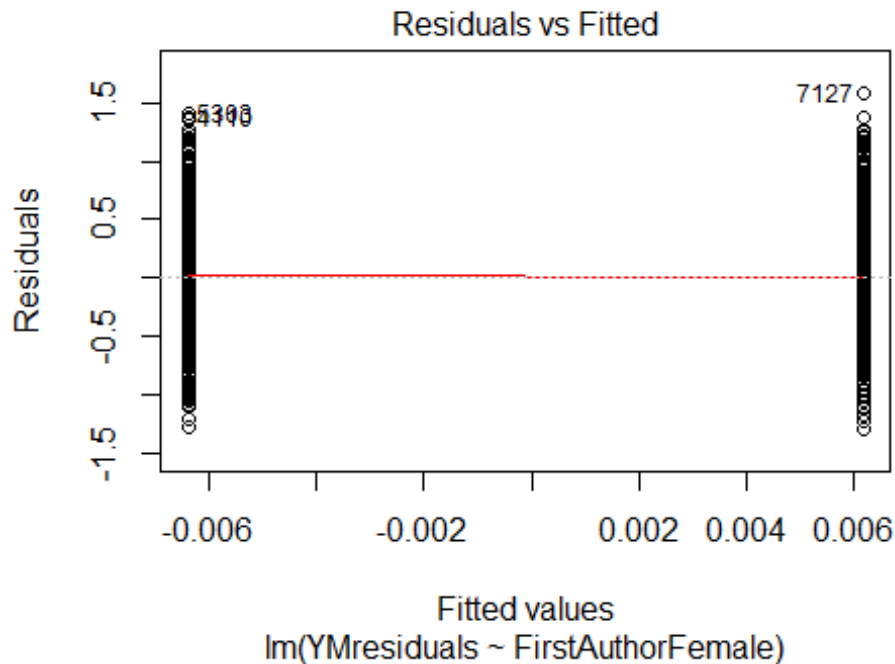
```

```
## 139 122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 92 83 68 66 53 73 70 102 76 114 130 117 99 91
## 2011 2012
## 126 100
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



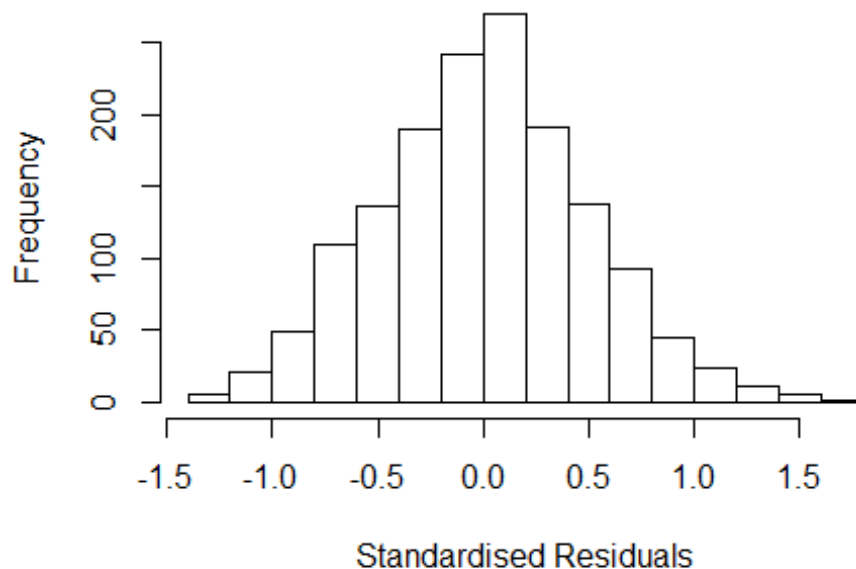
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.047, df = 1, p-value = 0.8
```





```
## [1] "Female first author team size 2018 geometric mean: 3.51003407066612"
## [1] "Male first author team size 2018 geometric mean: 3.69990721497157"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.51528267667893"
## [1] "Male last author team size 2018 geometric mean: 3.68096328848164"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1600, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.173 1      1.083
## LastAuthorFemale  1.186 1      1.089
## UniqueAuthors    1.312 4      1.035
## Year             1.426 16      1.011
```

## Residuals from first and last author and team size



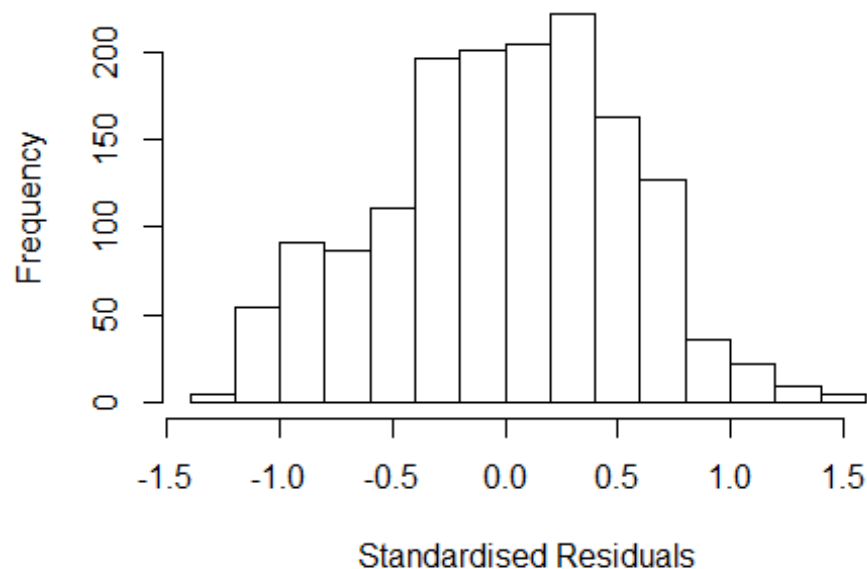
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3874 -0.3287 0.0116 0.3302 1.6209
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9038 0.0844 10.71 < 2e-16 ***
## FirstAuthorFemale1 -0.0411 0.0278 -1.48 0.14012
## LastAuthorFemale1 -0.0638 0.0298 -2.14 0.03252 *
## UniqueAuthors2 0.2204 0.0509 4.33 1.6e-05 ***
## UniqueAuthors3 0.3616 0.0504 7.17 1.1e-12 ***
## UniqueAuthors4 0.4440 0.0502 8.84 < 2e-16 ***
## UniqueAuthors5 0.6025 0.0442 13.63 < 2e-16 ***
## Year1997 -0.1210 0.0900 -1.34 0.17898
## Year1998 -0.1228 0.0881 -1.39 0.16342
## Year1999 -0.1173 0.0877 -1.34 0.18122
```

```

## Year2000          0.0369      0.0908      0.41  0.68476
## Year2001         -0.1406      0.1003     -1.40  0.16104
## Year2002         -0.2010      0.0918     -2.19  0.02867 *
## Year2003         -0.1301      0.0895     -1.45  0.14634
## Year2004         -0.2123      0.0836     -2.54  0.01115 *
## Year2005         -0.2166      0.0952     -2.28  0.02299 *
## Year2006         -0.3513      0.0840     -4.18  3.1e-05 ***
## Year2007         -0.2249      0.0886     -2.54  0.01121 *
## Year2008         -0.3170      0.0888     -3.57  0.00037 ***
## Year2009         -0.2003      0.0840     -2.38  0.01722 *
## Year2010         -0.2744      0.0876     -3.13  0.00176 **
## Year2011         -0.1189      0.0815     -1.46  0.14502
## Year2012         -0.2689      0.0918     -2.93  0.00344 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.185, Adjusted R-squared:  0.173
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 147 weights are ~= 1. The remaining 1384 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.859  0.949  0.903  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.116 1      1.056
## LastAuthorFemale  1.114 1      1.056
## Year              1.105 16      1.003

```

## Residuals from first and last author



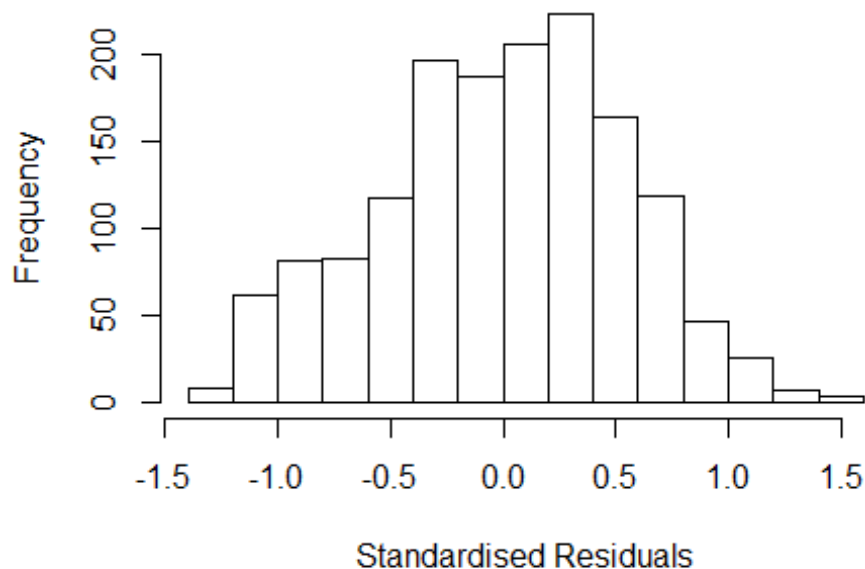
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3296 -0.3603 0.0159 0.3742 1.4889
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2909 0.0703 18.37 < 2e-16 ***
## FirstAuthorFemale1 -0.0110 0.0300 -0.37 0.71339
## LastAuthorFemale1 -0.1105 0.0329 -3.36 0.00081 ***
## Year1997 -0.2208 0.0935 -2.36 0.01838 *
## Year1998 -0.1829 0.0886 -2.06 0.03922 *
## Year1999 -0.1612 0.0915 -1.76 0.07833 .
## Year2000 0.0387 0.0916 0.42 0.67264
## Year2001 -0.1262 0.0961 -1.31 0.18951
## Year2002 -0.2701 0.0991 -2.73 0.00650 **
## Year2003 -0.1530 0.0919 -1.67 0.09604 .
## Year2004 -0.2122 0.0875 -2.42 0.01545 *
## Year2005 -0.2382 0.0995 -2.40 0.01673 *
```

```

## Year2006          -0.3876      0.0859   -4.51  6.9e-06 ***
## Year2007          -0.2324      0.0898   -2.59  0.00978 **
## Year2008          -0.3559      0.0905   -3.93  8.8e-05 ***
## Year2009          -0.2359      0.0853   -2.77  0.00576 **
## Year2010          -0.3358      0.0896   -3.75  0.00018 ***
## Year2011          -0.1452      0.0828   -1.75  0.07979 .
## Year2012          -0.3129      0.0962   -3.25  0.00117 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.546
## Multiple R-squared:  0.0478, Adjusted R-squared:  0.0365
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1397 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.436  0.870   0.950   0.910   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.054 1          1.027
## Year              1.054 16          1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3132 -0.3554 0.0239 0.3771 1.4362
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2623 0.0706 17.87 < 2e-16 ***
## FirstAuthorFemale1 -0.0371 0.0293 -1.27 0.20566
## Year1997 -0.2024 0.0946 -2.14 0.03256 *
## Year1998 -0.1804 0.0896 -2.01 0.04425 *
## Year1999 -0.1491 0.0930 -1.60 0.10907
## Year2000 0.0509 0.0923 0.55 0.58101
## Year2001 -0.1173 0.0968 -1.21 0.22565
## Year2002 -0.2631 0.0995 -2.64 0.00828 **
## Year2003 -0.1499 0.0926 -1.62 0.10557
## Year2004 -0.2034 0.0883 -2.30 0.02146 *
## Year2005 -0.2413 0.0999 -2.41 0.01587 *
## Year2006 -0.3786 0.0867 -4.36 1.4e-05 ***
```

```

## Year2007          -0.2226      0.0898   -2.48  0.01331 *
## Year2008          -0.3564      0.0912   -3.91  9.7e-05 ***
## Year2009          -0.2212      0.0861   -2.57  0.01030 *
## Year2010          -0.3365      0.0898   -3.75  0.00019 ***
## Year2011          -0.1411      0.0840   -1.68  0.09340 .
## Year2012          -0.3216      0.0976   -3.29  0.00101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.554
## Multiple R-squared:  0.0394, Adjusted R-squared:  0.0286
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1411 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.482  0.874  0.952  0.913  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.053 1      1.026
## Year      1.053 16      1.002

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.3263 -0.3610  0.0179  0.3737  1.4855

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2873    0.0698   18.43 < 2e-16 ***
## LastAuthorFemale1 -0.1135    0.0320   -3.54 0.00040 ***
## Year1997         -0.2211    0.0935   -2.36 0.01817 *
## Year1998         -0.1833    0.0886   -2.07 0.03882 *
## Year1999         -0.1620    0.0914   -1.77 0.07647 .
## Year2000          0.0390    0.0917    0.43 0.67055
## Year2001         -0.1270    0.0961   -1.32 0.18673
## Year2002         -0.2697    0.0993   -2.72 0.00669 **
## Year2003         -0.1540    0.0920   -1.67 0.09415 .
## Year2004         -0.2128    0.0875   -2.43 0.01513 *
## Year2005         -0.2392    0.0994   -2.41 0.01628 *
## Year2006         -0.3886    0.0858   -4.53 6.4e-06 ***
## Year2007         -0.2342    0.0895   -2.62 0.00900 **
## Year2008         -0.3568    0.0905   -3.94 8.4e-05 ***
## Year2009         -0.2372    0.0854   -2.78 0.00553 **
## Year2010         -0.3375    0.0894   -3.78 0.00016 ***
## Year2011         -0.1462    0.0827   -1.77 0.07741 .
## Year2012         -0.3143    0.0960   -3.27 0.00109 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.546
## Multiple R-squared:  0.0477, Adjusted R-squared:  0.037
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1397 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.439  0.871  0.950  0.910  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1531"
## [1] ""

```



```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2709"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 14 19 15 21 11 10 2 2 14 13 17 7 17 20
## 2011 2012
## 10 8
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 9 18 12 14 10 10 1 2 14 12 13 6 17 18
## 2011 2012
## 10 7
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 8 18 11 14 10 10 1 2 14 12 13 6 17 18
## 2011 2012
## 9 7
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.91947121957741"
## [1] "Male first author team size 2018 geometric mean: 1.31607401295249"

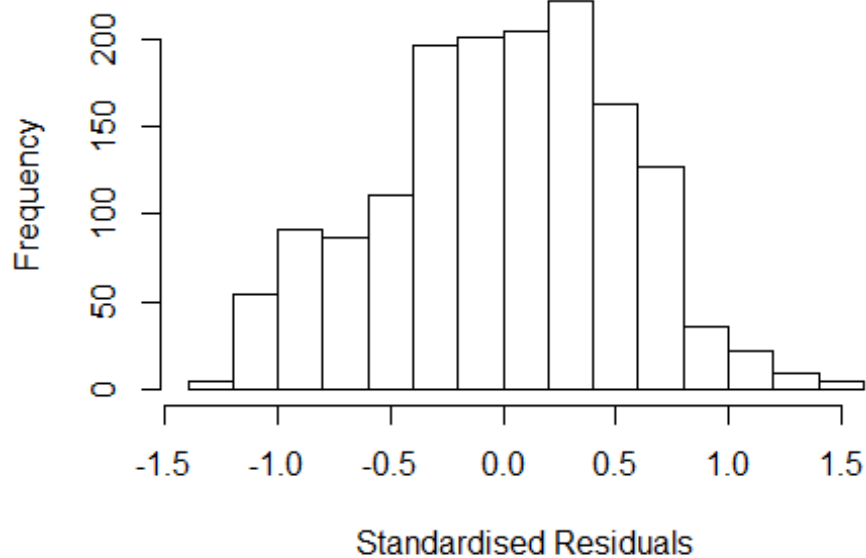
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.9740317478448"
## [1] "Male last author team size 2018 geometric mean: 1.31950791077289"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

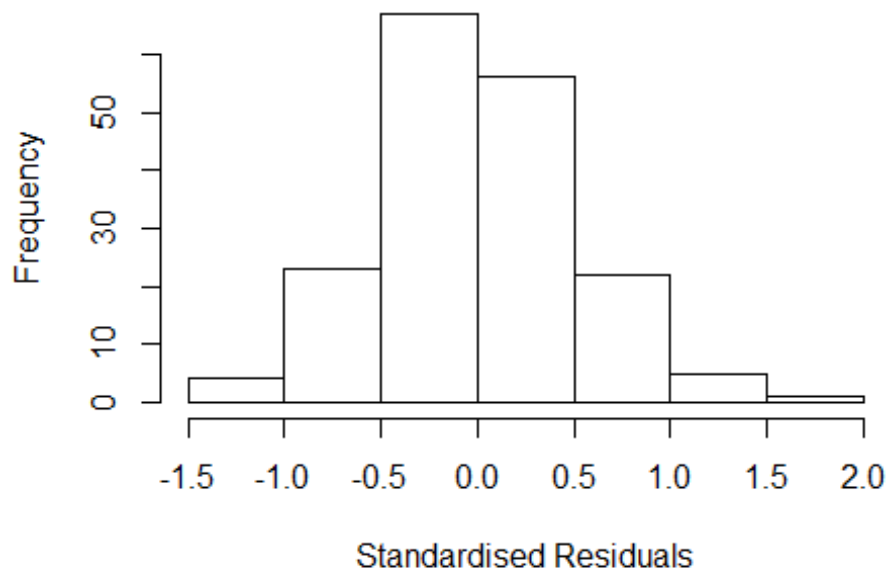
```

## Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 46, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 7.313e+00 1          2.704
## LastAuthorFemale -1.831e+14 1           NaN
## UniqueAuthors    -4.880e+15 4           NaN
## Year              -9.686e+15 16           NaN
```

## Residuals from first and last author and team size



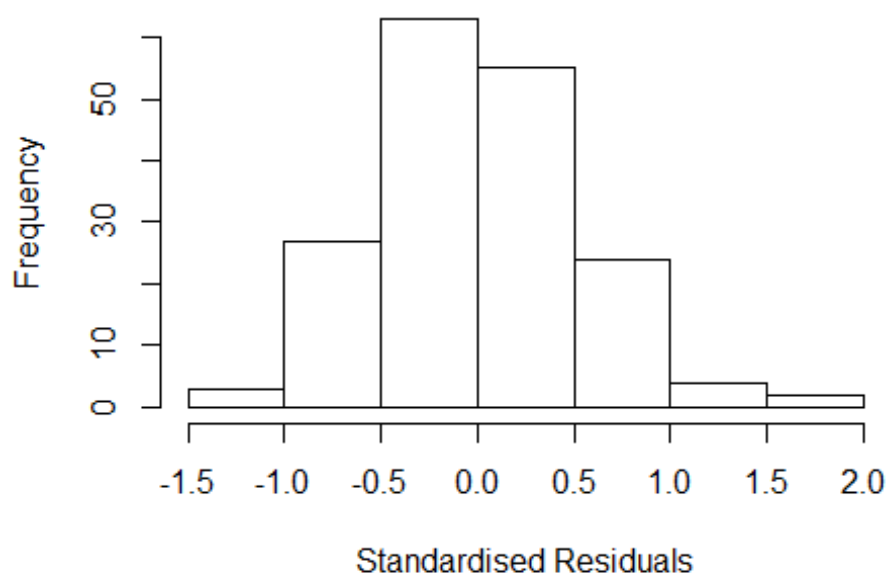
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.115 -0.361 -0.040 0.351 1.686
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.68583 0.32745 2.09 0.0378 *
## FirstAuthorFemale1 0.34124 0.10675 3.20 0.0017 **
## LastAuthorFemale1 -0.47256 0.10951 -4.32 2.8e-05 ***
## UniqueAuthors2 -0.06724 0.09181 -0.73 0.4650
## UniqueAuthors3 -0.00135 0.18150 -0.01 0.9941
## UniqueAuthors4 0.85984 0.18381 4.68 6.3e-06 ***
## UniqueAuthors5 0.11416 0.20801 0.55 0.5839
## Year1997 0.29739 0.38949 0.76 0.4463
## Year1998 -0.09881 0.34434 -0.29 0.7745
## Year1999 0.08856 0.36208 0.24 0.8071
```

```

## Year2000          0.37554    0.36382    1.03    0.3036
## Year2001          0.71369    0.35711    2.00    0.0474 *
## Year2002          0.18417    0.35841    0.51    0.6081
## Year2003          0.52917    0.32745    1.62    0.1081
## Year2004         -0.03327    0.36313   -0.09    0.9271
## Year2005          0.32683    0.34626    0.94    0.3467
## Year2006          0.32891    0.34353    0.96    0.3398
## Year2007          0.27859    0.35576    0.78    0.4348
## Year2008          0.39421    0.38428    1.03    0.3066
## Year2009          0.37103    0.36838    1.01    0.3154
## Year2010          0.24388    0.35814    0.68    0.4969
## Year2011          0.80595    0.36751    2.19    0.0298 *
## Year2012          0.42927    0.38335    1.12    0.2645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.217, Adjusted R-squared:  0.106
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 168 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.282  0.900  0.957  0.919  0.987  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 5.457 1      2.336
## LastAuthorFemale  4.622 1      2.150
## Year              6.357 16      1.060

```

## Residuals from first and last author



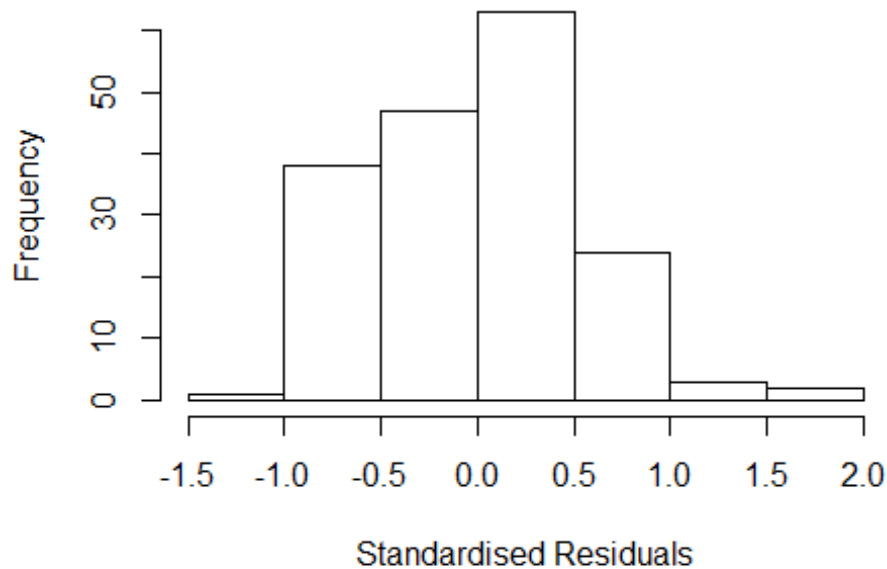
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1124 -0.3785 -0.0589 0.3626 1.6808
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.65394 0.32238 2.03 0.0442 *
## FirstAuthorFemale1 0.32331 0.10884 2.97 0.0034 **
## LastAuthorFemale1 -0.45510 0.10768 -4.23 4e-05 ***
## Year1997 0.32435 0.38958 0.83 0.4063
## Year1998 -0.06178 0.34548 -0.18 0.8583
## Year1999 0.11185 0.36249 0.31 0.7581
## Year2000 0.40755 0.36031 1.13 0.2597
## Year2001 0.72096 0.36672 1.97 0.0510 .
## Year2002 0.20881 0.36191 0.58 0.5648
## Year2003 0.56106 0.32238 1.74 0.0837 .
## Year2004 0.00691 0.34951 0.02 0.9843
## Year2005 0.33926 0.35093 0.97 0.3351
```

```

## Year2006          0.34032      0.34882      0.98      0.3307
## Year2007          0.29741      0.35893      0.83      0.4086
## Year2008          0.41248      0.38242      1.08      0.2824
## Year2009          0.39597      0.37498      1.06      0.2926
## Year2010          0.34522      0.35844      0.96      0.3370
## Year2011          0.83142      0.36601      2.27      0.0245 *
## Year2012          0.45842      0.38103      1.20      0.2307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.202, Adjusted R-squared:  0.112
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 172 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.283  0.904  0.953  0.919  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.007 1      1.734
## Year              3.007 16      1.035

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0186 -0.3943 0.0547 0.3527 1.6780
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5861 0.3388 1.73 0.086 .
## FirstAuthorFemale1 -0.0176 0.0880 -0.20 0.842
## Year1997 0.4176 0.3891 1.07 0.285
## Year1998 0.0089 0.3647 0.02 0.981
## Year1999 0.1524 0.3805 0.40 0.689
## Year2000 0.4325 0.3695 1.17 0.244
## Year2001 0.6847 0.3843 1.78 0.077 .
## Year2002 0.2081 0.3795 0.55 0.584
## Year2003 0.6289 0.3388 1.86 0.065 .
## Year2004 0.2452 0.4108 0.60 0.551
## Year2005 0.3324 0.3682 0.90 0.368
## Year2006 0.2719 0.3709 0.73 0.465
```

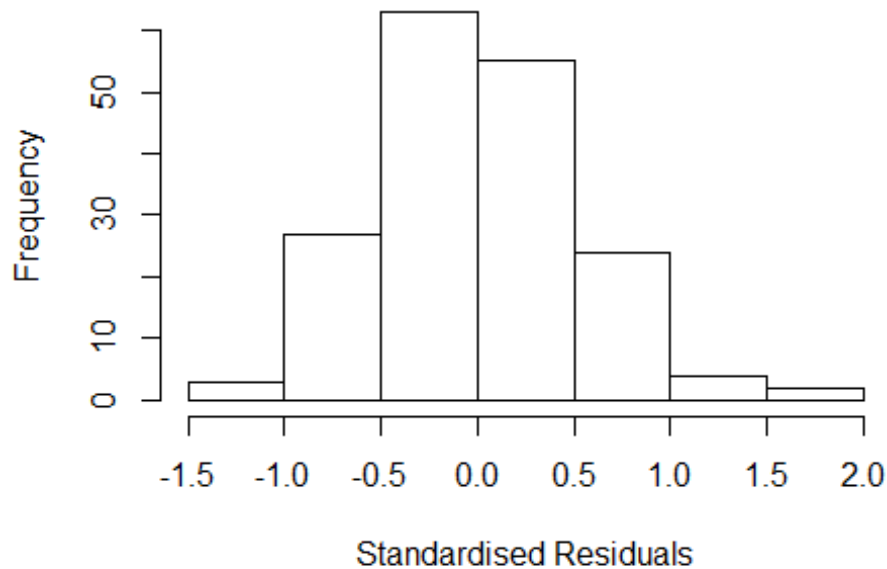
```

## Year2007          0.3918      0.3740      1.05      0.296
## Year2008          0.4418      0.3981      1.11      0.269
## Year2009          0.3555      0.3897      0.91      0.363
## Year2010          0.3785      0.3836      0.99      0.325
## Year2011          0.8727      0.3853      2.26      0.025 *
## Year2012          0.3871      0.4127      0.94      0.350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.13,   Adjusted R-squared:  0.0381
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 169 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.369  0.909   0.957   0.924   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 2.695 1      1.642
## Year              2.695 16      1.031

```



## Residuals from last author



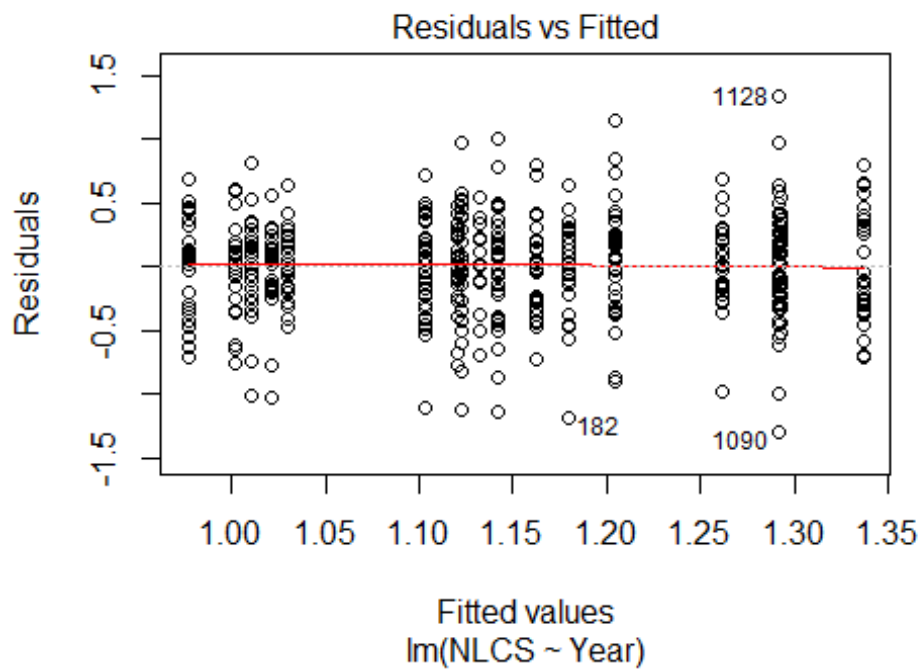
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1120 -0.4277 0.0126 0.3688 1.5916
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6354 0.3311 1.92 0.057 .
## LastAuthorFemale1 -0.2166 0.0874 -2.48 0.014 *
## Year1997 0.3952 0.3817 1.04 0.302
## Year1998 0.0460 0.3454 0.13 0.894
## Year1999 0.1640 0.3669 0.45 0.656
## Year2000 0.4549 0.3693 1.23 0.220
## Year2001 0.7629 0.3854 1.98 0.049 *
## Year2002 0.2800 0.3645 0.77 0.444
## Year2003 0.5796 0.3311 1.75 0.082 .
## Year2004 0.1871 0.3962 0.47 0.637
## Year2005 0.4000 0.3641 1.10 0.274
## Year2006 0.3503 0.3608 0.97 0.333
```

```

## Year2007          0.4481      0.3572      1.25      0.211
## Year2008          0.4583      0.3888      1.18      0.240
## Year2009          0.4488      0.3841      1.17      0.244
## Year2010          0.4220      0.3695      1.14      0.255
## Year2011          0.9444      0.3681      2.57      0.011 *
## Year2012          0.4766      0.4021      1.19      0.238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.168, Adjusted R-squared:  0.0792
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 168 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.375  0.906   0.947   0.920   0.984   0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 178"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2710"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   67   68   68   76   93   88   76   42   40   31   51   41   36   54   52
## 2011 2012
##   57   55
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   29   25   24   41   23   27   26   28   16   35   29   24   34   29
## 2011 2012

```

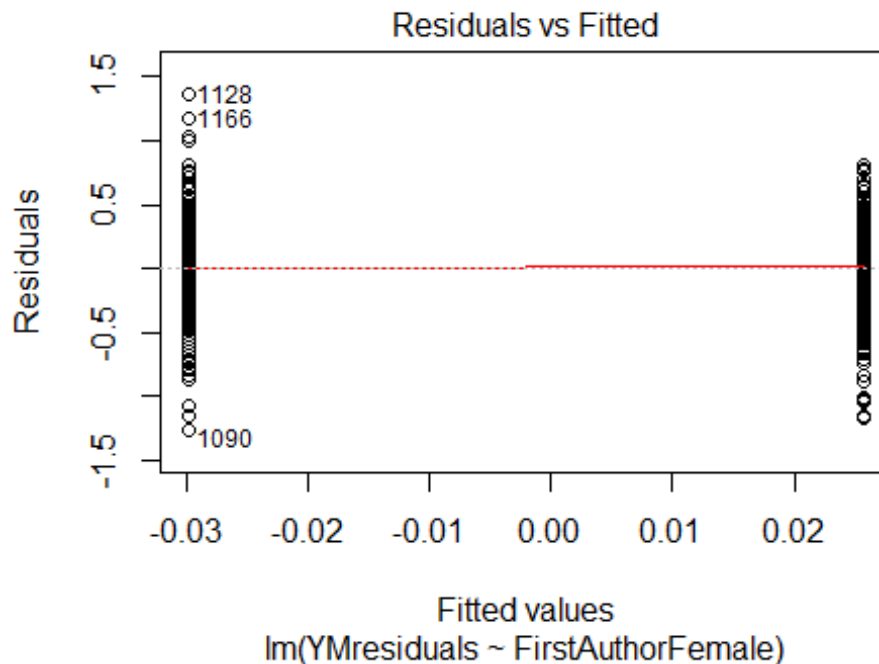
```
## 44 37
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 23 26 22 22 36 23 23 21 22 14 32 28 17 27 20
## 2011 2012
## 42 34
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
## [1] "Female first author team size 2018 geometric mean: 3.22036453952603"
## [1] "Male first author team size 2018 geometric mean: 1.77051949862246"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

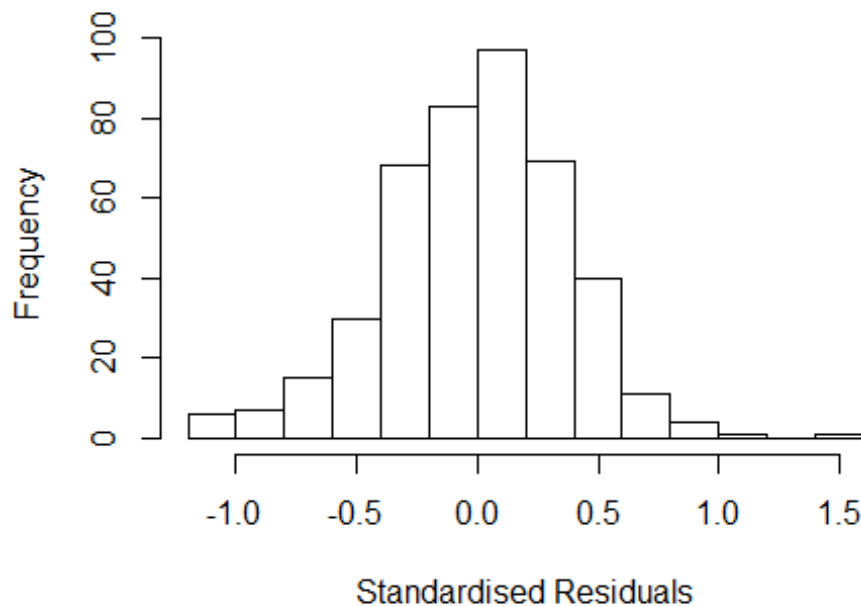
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.83552791881973"
## [1] "Male last author team size 2018 geometric mean: 2.1965624705903"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.552 1          1.246
## LastAuthorFemale  1.456 1          1.207
## UniqueAuthors    2.177 4          1.102
## Year              3.009 16          1.035
```

## Residuals from first and last author and team size



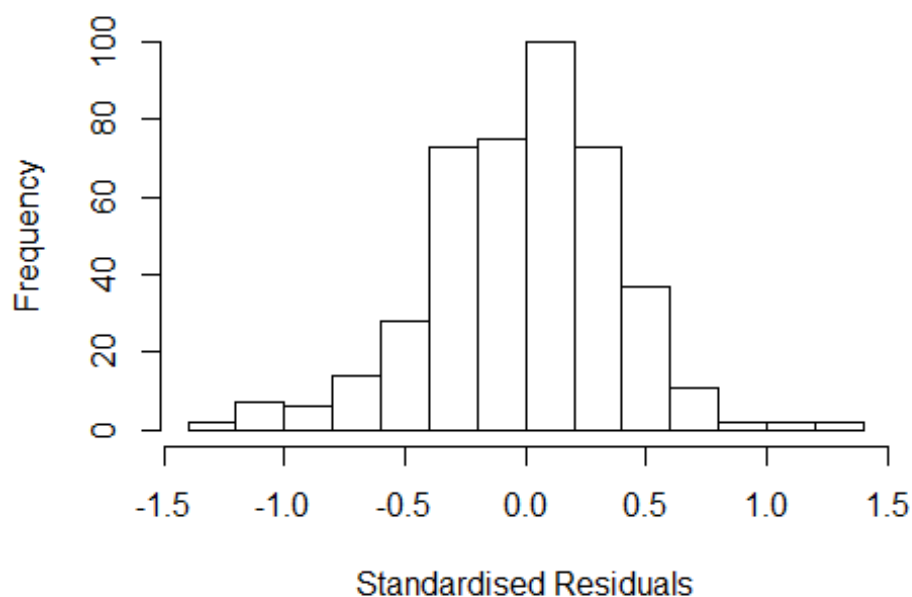
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.13910 -0.25413  0.00839  0.24772  1.57682
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07890    0.10482   10.29 < 2e-16 ***
## FirstAuthorFemale1  0.03678    0.04368    0.84  0.40025
## LastAuthorFemale1  0.09027    0.04511    2.00  0.04604 *
## UniqueAuthors2     0.19090    0.10634    1.80  0.07335 .
## UniqueAuthors3     0.18609    0.10552    1.76  0.07857 .
## UniqueAuthors4     0.14125    0.10829    1.30  0.19283
## UniqueAuthors5     0.24583    0.10272    2.39  0.01715 *
## Year1997          -0.12780    0.09464   -1.35  0.17764
## Year1998          -0.11376    0.08900   -1.28  0.20188
## Year1999          -0.25037    0.08996   -2.78  0.00563 **
```

```

## Year2000      -0.28068    0.08731   -3.21  0.00141 **
## Year2001      -0.31290    0.08148   -3.84  0.00014 ***
## Year2002      -0.17819    0.08315   -2.14  0.03270 *
## Year2003      -0.27866    0.08555   -3.26  0.00122 **
## Year2004      -0.29190    0.10398   -2.81  0.00524 **
## Year2005      -0.23054    0.11641   -1.98  0.04833 *
## Year2006      -0.17033    0.08219   -2.07  0.03886 *
## Year2007      -0.16749    0.11764   -1.42  0.15527
## Year2008      -0.08503    0.09601   -0.89  0.37636
## Year2009      -0.17101    0.11021   -1.55  0.12153
## Year2010      -0.00409    0.12697   -0.03  0.97434
## Year2011      -0.03172    0.09045   -0.35  0.72600
## Year2012      -0.14376    0.10780   -1.33  0.18306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.368
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0629
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 392 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0262 0.8810 0.9490 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.31e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.329 1          1.153
## LastAuthorFemale  1.407 1          1.186
## Year              1.757 16          1.018

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2410 -0.2562 0.0251 0.2341 1.3830
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25668 0.06129 20.50 < 2e-16 ***
## FirstAuthorFemale1 0.05636 0.04099 1.38 0.16987
## LastAuthorFemale1 0.09632 0.04501 2.14 0.03297 *
## Year1997 -0.14555 0.09563 -1.52 0.12877
## Year1998 -0.13029 0.09288 -1.40 0.16144
## Year1999 -0.25322 0.08705 -2.91 0.00382 **
## Year2000 -0.30729 0.08733 -3.52 0.00048 ***
## Year2001 -0.30372 0.08074 -3.76 0.00019 ***
## Year2002 -0.18514 0.08318 -2.23 0.02657 *
## Year2003 -0.26904 0.08317 -3.23 0.00131 **
## Year2004 -0.30124 0.10426 -2.89 0.00406 **
## Year2005 -0.21361 0.11488 -1.86 0.06368 .
```

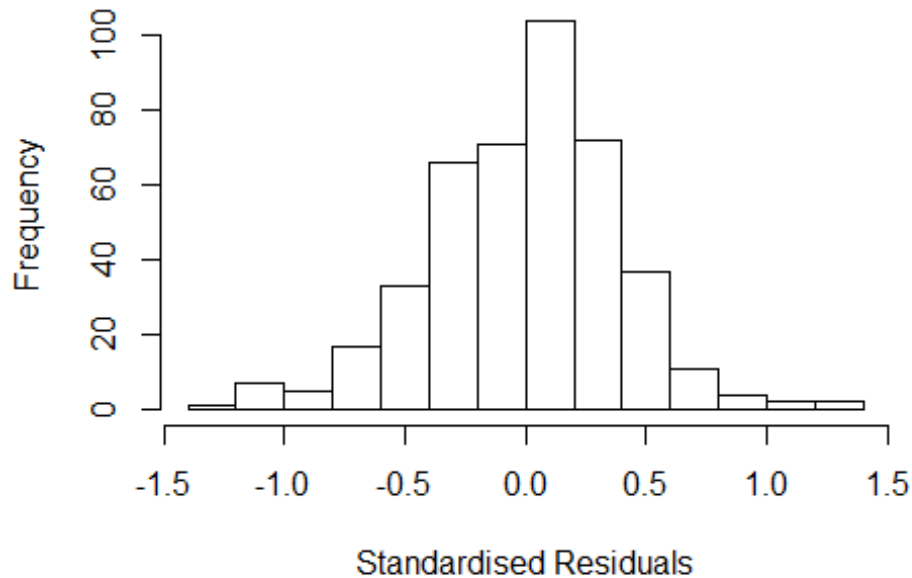
```

## Year2006      -0.17416    0.08452   -2.06  0.03996 *
## Year2007      -0.18480    0.11192   -1.65  0.09946 .
## Year2008      -0.09785    0.09434   -1.04  0.30025
## Year2009      -0.17115    0.10784   -1.59  0.11327
## Year2010       0.00611    0.11729    0.05  0.95851
## Year2011      -0.01565    0.08762   -0.18  0.85831
## Year2012      -0.14826    0.10698   -1.39  0.16653
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0871, Adjusted R-squared:  0.0473
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~= 1. The remaining 393 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.131  0.878  0.951  0.900  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.284 1      1.133
## Year              1.284 16      1.008

```



## Residuals from first author



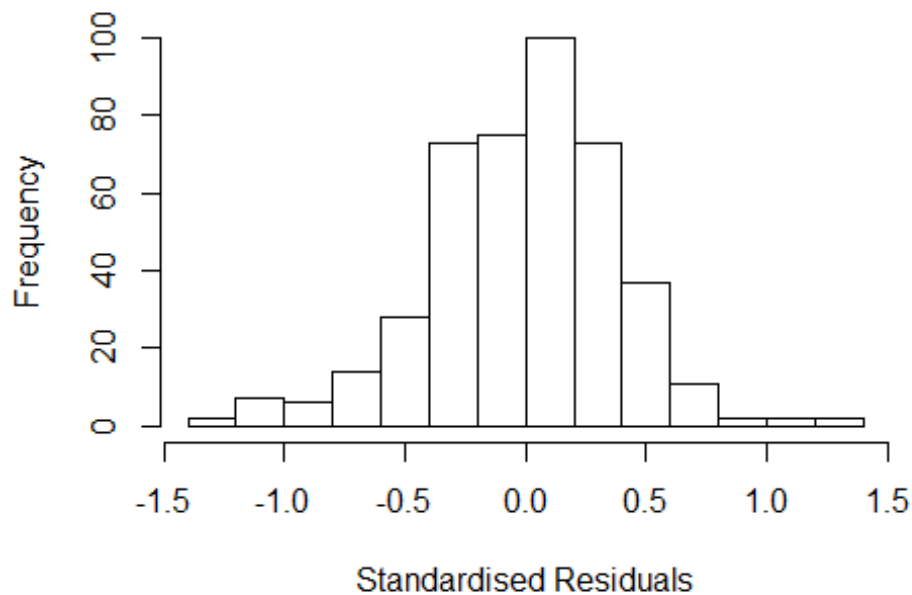
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2473 -0.2538 0.0212 0.2285 1.3767
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.27852 0.06373 20.06 < 2e-16 ***
## FirstAuthorFemale1 0.07236 0.04083 1.77 0.07707 .
## Year1997 -0.14561 0.09727 -1.50 0.13519
## Year1998 -0.12585 0.09456 -1.33 0.18393
## Year1999 -0.26551 0.09044 -2.94 0.00351 **
## Year2000 -0.31220 0.08833 -3.53 0.00045 ***
## Year2001 -0.28241 0.08167 -3.46 0.00060 ***
## Year2002 -0.17527 0.08740 -2.01 0.04556 *
## Year2003 -0.27104 0.08693 -3.12 0.00195 **
## Year2004 -0.30677 0.10351 -2.96 0.00321 **
## Year2005 -0.15567 0.11369 -1.37 0.17166
## Year2006 -0.17954 0.08750 -2.05 0.04080 *
```

```

## Year2007          -0.19003      0.11653    -1.63   0.10371
## Year2008          -0.10784      0.09812    -1.10   0.27240
## Year2009          -0.19579      0.11019    -1.78   0.07633 .
## Year2010           0.00299      0.12024      0.02   0.98019
## Year2011          -0.03118      0.09258    -0.34   0.73644
## Year2012          -0.15049      0.10973    -1.37   0.17095
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.373
## Multiple R-squared:  0.0742, Adjusted R-squared:  0.0362
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 36 weights are ~= 1. The remaining 396 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.143  0.883  0.952  0.901  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.359 1          1.166
## Year            1.359 16          1.010

```

## Residuals from last author



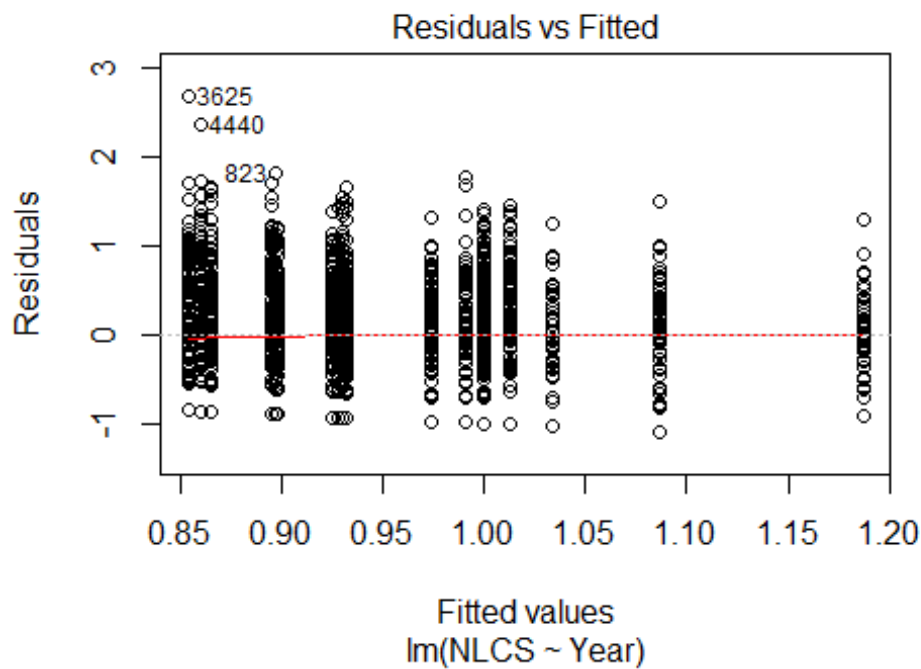
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2746 -0.2620 0.0255 0.2263 1.3494
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.28037 0.05852 21.88 < 2e-16 ***
## LastAuthorFemale1 0.10793 0.04445 2.43 0.01559 *
## Year1997 -0.15392 0.09502 -1.62 0.10602
## Year1998 -0.11879 0.09337 -1.27 0.20398
## Year1999 -0.24573 0.08740 -2.81 0.00516 **
## Year2000 -0.30784 0.08794 -3.50 0.00051 ***
## Year2001 -0.31093 0.08219 -3.78 0.00018 ***
## Year2002 -0.18531 0.08310 -2.23 0.02629 *
## Year2003 -0.26155 0.08341 -3.14 0.00184 **
## Year2004 -0.29668 0.10450 -2.84 0.00475 **
## Year2005 -0.21998 0.11327 -1.94 0.05281 .
## Year2006 -0.16881 0.08508 -1.98 0.04792 *
```

```

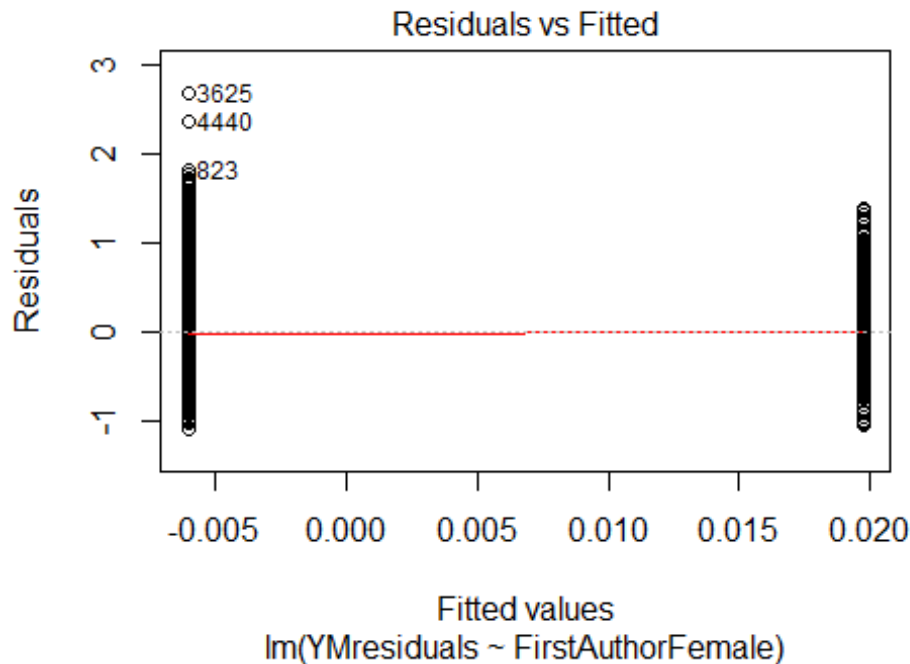
## Year2007          -0.17640      0.11150    -1.58  0.11438
## Year2008          -0.10185      0.09370    -1.09  0.27764
## Year2009          -0.17040      0.10711    -1.59  0.11241
## Year2010           0.01709      0.11848      0.14  0.88540
## Year2011          -0.00575      0.08694    -0.07  0.94727
## Year2012          -0.12749      0.10330    -1.23  0.21786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.37
## Multiple R-squared:  0.0828, Adjusted R-squared:  0.0451
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 394 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.154  0.879  0.948  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 432"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2711"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 260 230 240 200 223 266 259 245 286 351 381 407 342 310 295
## 2011 2012
## 263 236
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 50 45 54 65 58 108 100 114 129 163 178 225 171 127 183
## 2011 2012

```

```
## 165 159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 33 46 59 52 83 74 79 82 94 119 180 118 100 143
## 2011 2012
## 125 128
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```

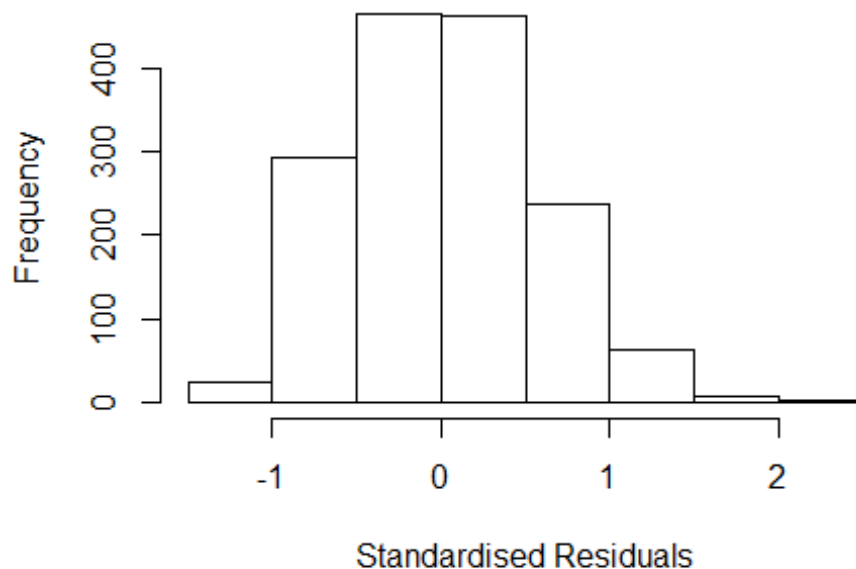


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.5, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 3.17071644212789"
## [1] "Male first author team size 2018 geometric mean: 3.26375489678893"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 810, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.40075835281588"
## [1] "Male last author team size 2018 geometric mean: 3.2066515112925"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 830, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.145 1      1.070
## LastAuthorFemale  1.158 1      1.076
## UniqueAuthors    1.389 4      1.042
## Year             1.530 16      1.013
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.48641 -0.42131 -0.00432 0.39443 2.26516
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9584 0.0740 12.95 < 2e-16 ***
## FirstAuthorFemale1 0.0382 0.0354 1.08 0.28065
## LastAuthorFemale1 -0.0153 0.0392 -0.39 0.69616
## UniqueAuthors2 0.1193 0.0455 2.62 0.00888 **
## UniqueAuthors3 0.2194 0.0465 4.72 2.6e-06 ***
## UniqueAuthors4 0.3832 0.0522 7.34 3.5e-13 ***
## UniqueAuthors5 0.6381 0.0545 11.71 < 2e-16 ***
## Year1997 -0.1101 0.1226 -0.90 0.36955
## Year1998 -0.0568 0.1102 -0.52 0.60653
## Year1999 -0.1942 0.1030 -1.89 0.05951 .
```

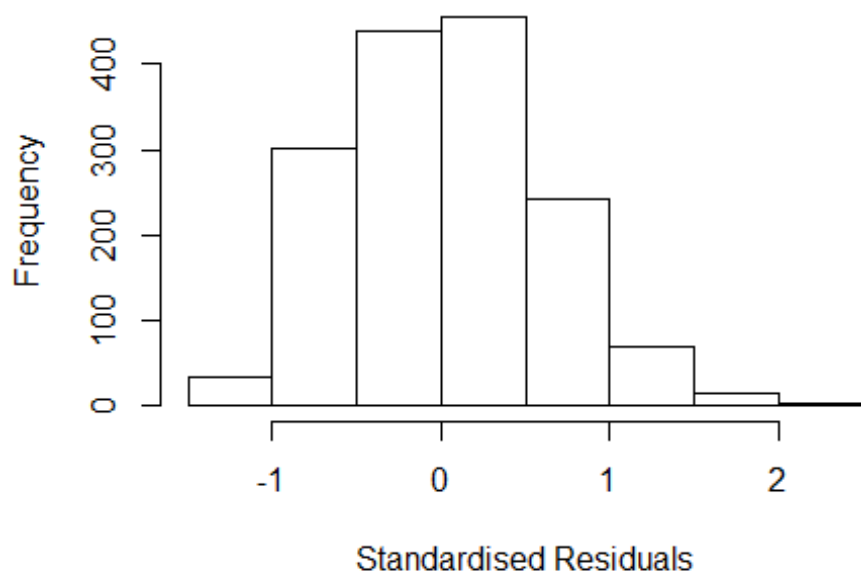
```

## Year2000          -0.1643      0.1066    -1.54    0.12334
## Year2001          -0.2122      0.0926    -2.29    0.02199 *
## Year2002          -0.1829      0.0999    -1.83    0.06728 .
## Year2003          -0.2323      0.0902    -2.58    0.01010 *
## Year2004          -0.3135      0.0923    -3.40    0.00070 ***
## Year2005          -0.2938      0.0935    -3.14    0.00172 **
## Year2006          -0.3802      0.0933    -4.07    4.9e-05 ***
## Year2007          -0.2070      0.0843    -2.45    0.01422 *
## Year2008          -0.4659      0.0879    -5.30    1.3e-07 ***
## Year2009          -0.3622      0.0967    -3.74    0.00019 ***
## Year2010          -0.3958      0.0894    -4.43    1.0e-05 ***
## Year2011          -0.3129      0.0889    -3.52    0.00044 ***
## Year2012          -0.2780      0.0905    -3.07    0.00218 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.597
## Multiple R-squared:  0.125, Adjusted R-squared:  0.113
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1439 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.118  0.875   0.950   0.916   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.43e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.158 1          1.076
## LastAuthorFemale  1.137 1          1.066
## Year              1.142 16          1.004

```



## Residuals from first and last author



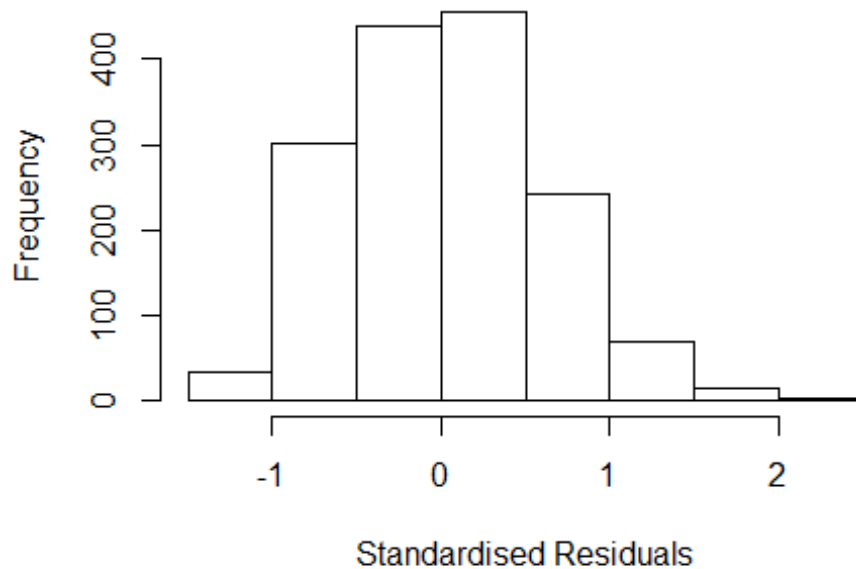
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10333 -0.41404 0.00829 0.42401 2.33633
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1352 0.0660 17.20 < 2e-16 ***
## FirstAuthorFemale1 0.0366 0.0383 0.96 0.3394
## LastAuthorFemale1 -0.0135 0.0420 -0.32 0.7485
## Year1997 -0.0723 0.1202 -0.60 0.5479
## Year1998 -0.0319 0.1114 -0.29 0.7749
## Year1999 -0.2696 0.1051 -2.56 0.0104 *
## Year2000 -0.2095 0.1060 -1.98 0.0483 *
## Year2001 -0.2045 0.0921 -2.22 0.0266 *
## Year2002 -0.1654 0.0956 -1.73 0.0839 .
## Year2003 -0.1854 0.0907 -2.04 0.0411 *
## Year2004 -0.2908 0.0897 -3.24 0.0012 **
## Year2005 -0.2647 0.0909 -2.91 0.0036 **
```

```

## Year2006          -0.3548      0.0907    -3.91  9.5e-05 ***
## Year2007          -0.1357      0.0819    -1.66  0.0980 .
## Year2008          -0.3954      0.0887    -4.46  8.8e-06 ***
## Year2009          -0.2482      0.0977    -2.54  0.0112 *
## Year2010          -0.2605      0.0885    -2.94  0.0033 **
## Year2011          -0.1832      0.0869    -2.11  0.0351 *
## Year2012          -0.1009      0.0909    -1.11  0.2674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.626
## Multiple R-squared:  0.0256, Adjusted R-squared:  0.0141
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1432 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.133  0.869  0.950  0.914  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1      1.041
## Year              1.084 16      1.003

```

## Residuals from first author



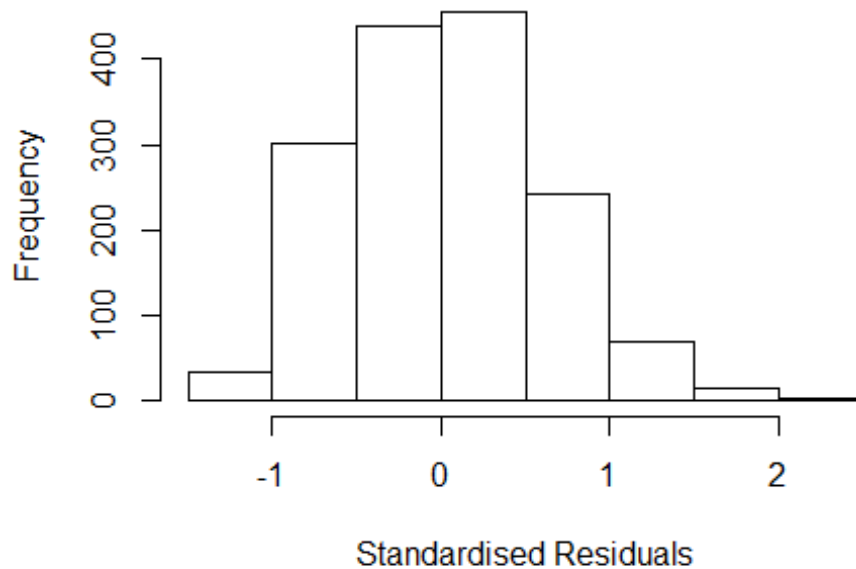
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10072 -0.41289 0.00617 0.42522 2.33748
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1349 0.0659 17.23 < 2e-16 ***
## FirstAuthorFemale1 0.0334 0.0371 0.90 0.3680
## Year1997 -0.0723 0.1201 -0.60 0.5471
## Year1998 -0.0342 0.1109 -0.31 0.7578
## Year1999 -0.2708 0.1049 -2.58 0.0100 **
## Year2000 -0.2121 0.1056 -2.01 0.0447 *
## Year2001 -0.2063 0.0920 -2.24 0.0250 *
## Year2002 -0.1663 0.0955 -1.74 0.0817 .
## Year2003 -0.1869 0.0904 -2.07 0.0389 *
## Year2004 -0.2920 0.0894 -3.27 0.0011 **
## Year2005 -0.2652 0.0908 -2.92 0.0035 **
## Year2006 -0.3568 0.0899 -3.97 7.5e-05 ***
```

```

## Year2007          -0.1375      0.0813   -1.69    0.0911 .
## Year2008          -0.3965      0.0884   -4.49    7.8e-06 ***
## Year2009          -0.2500      0.0977   -2.56    0.0106 *
## Year2010          -0.2614      0.0882   -2.96    0.0031 **
## Year2011          -0.1848      0.0866   -2.13    0.0331 *
## Year2012          -0.1021      0.0908   -1.12    0.2609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0255, Adjusted R-squared:  0.0147
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1428 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.135  0.869  0.950  0.914  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.066 1          1.032
## Year              1.066 16          1.002

```

## Residuals from last author



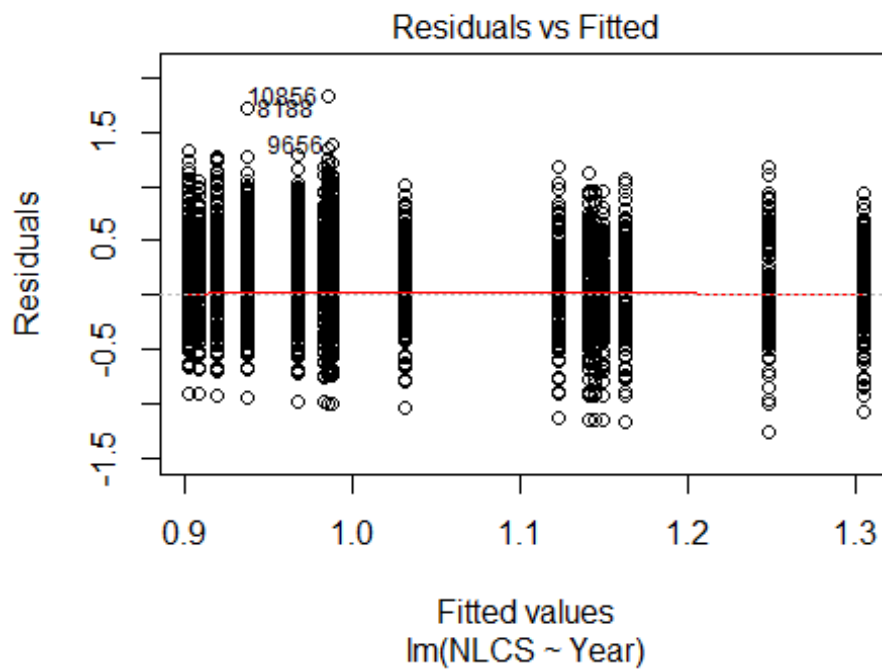
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.10839 -0.41682 0.00252 0.42995 2.32539
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14057 0.06571 17.36 < 2e-16 ***
## LastAuthorFemale1 -0.00302 0.04049 -0.07 0.9405
## Year1997 -0.07120 0.12010 -0.59 0.5534
## Year1998 -0.03218 0.11181 -0.29 0.7735
## Year1999 -0.26914 0.10532 -2.56 0.0107 *
## Year2000 -0.20393 0.10627 -1.92 0.0552 .
## Year2001 -0.20251 0.09196 -2.20 0.0278 *
## Year2002 -0.16132 0.09532 -1.69 0.0908 .
## Year2003 -0.18097 0.09047 -2.00 0.0456 *
## Year2004 -0.29401 0.08985 -3.27 0.0011 **
## Year2005 -0.26370 0.09069 -2.91 0.0037 **
## Year2006 -0.35520 0.09056 -3.92 9.2e-05 ***
```

```

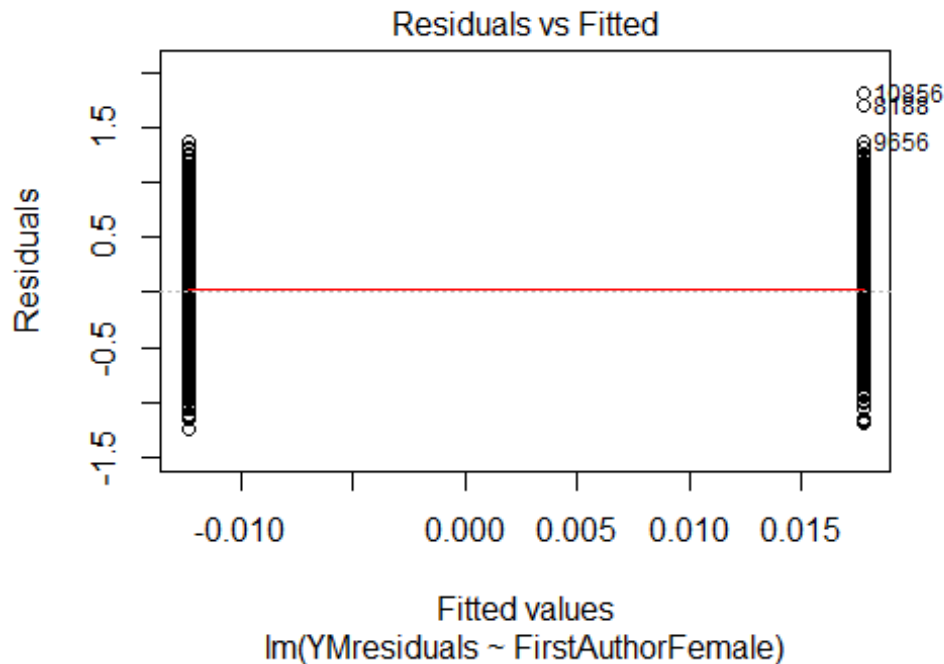
## Year2007          -0.13240      0.08178      -1.62      0.1056
## Year2008          -0.39487      0.08843      -4.47      8.6e-06 ***
## Year2009          -0.24893      0.09784      -2.54      0.0110 *
## Year2010          -0.25496      0.08802      -2.90      0.0038 **
## Year2011          -0.18346      0.08673      -2.12      0.0346 *
## Year2012          -0.09707      0.09073      -1.07      0.2848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.0143
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~ = 1. The remaining 1437 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.137  0.872  0.951  0.915  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.43e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1556"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2712"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  475  451  471  456  434  526  475  440  462  490  517  529  525  546  571
## 2011 2012
##  541  565
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  147  133  135  144  135  154  162  205  218  264  279  281  300  301  289
## 2011 2012

```

```
## 260 296
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 118 121 109 108 132 132 152 184 200 234 246 260 256 253
## 2011 2012
## 228 261
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 79, df = 16, p-value = 2e-10
```



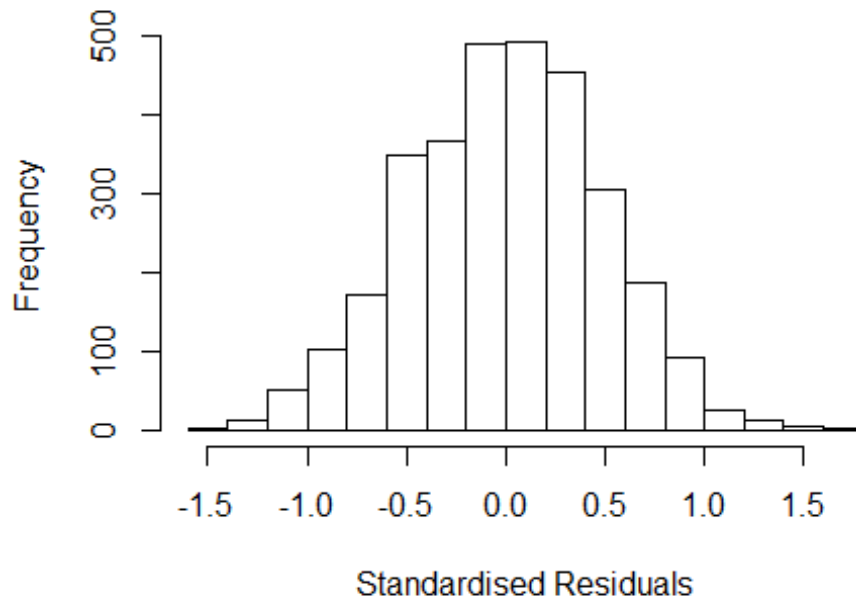
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.29, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 5.0694069329141"
## [1] "Male first author team size 2018 geometric mean: 4.66622774517232"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.98352437692809"
## [1] "Male last author team size 2018 geometric mean: 4.80247591669616"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5600, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.051 1 1.025
## LastAuthorFemale 1.070 1 1.035
## UniqueAuthors 1.184 4 1.021
## Year 1.214 16 1.006
```



## Residuals from first and last author and team size



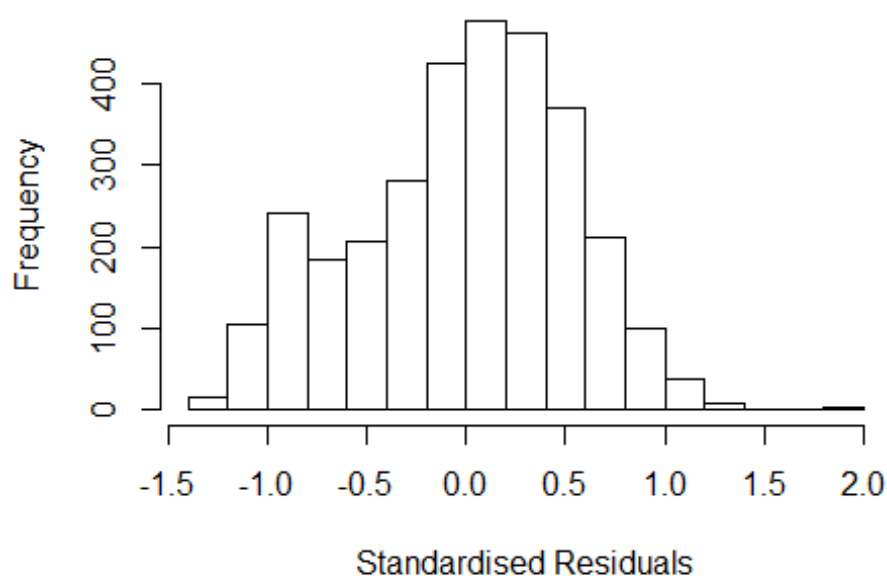
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.50472 -0.33985 0.00421 0.32735 1.61598
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.73657 0.06220 11.84 < 2e-16 ***
## FirstAuthorFemale1 -0.00102 0.01859 -0.05 0.95644
## LastAuthorFemale1 -0.01012 0.02188 -0.46 0.64383
## UniqueAuthors2 0.24866 0.04267 5.83 6.2e-09 ***
## UniqueAuthors3 0.36519 0.03978 9.18 < 2e-16 ***
## UniqueAuthors4 0.56949 0.03736 15.24 < 2e-16 ***
## UniqueAuthors5 0.71250 0.02986 23.86 < 2e-16 ***
## Year1997 0.06577 0.07236 0.91 0.36348
## Year1998 0.04398 0.06599 0.67 0.50516
## Year1999 -0.00593 0.07046 -0.08 0.93296
```

```

## Year2000      -0.07567      0.06784      -1.12      0.26474
## Year2001      -0.07481      0.06578      -1.14      0.25552
## Year2002      -0.10703      0.06935      -1.54      0.12288
## Year2003      -0.26555      0.06654      -3.99      6.7e-05 ***
## Year2004      -0.28938      0.06510      -4.45      9.1e-06 ***
## Year2005      -0.17470      0.06345      -2.75      0.00594 **
## Year2006      -0.27560      0.06388      -4.31      1.6e-05 ***
## Year2007      -0.22625      0.06339      -3.57      0.00036 ***
## Year2008      -0.28614      0.06429      -4.45      8.8e-06 ***
## Year2009      -0.25570      0.06409      -3.99      6.8e-05 ***
## Year2010      -0.26480      0.06694      -3.96      7.8e-05 ***
## Year2011      -0.20407      0.06565      -3.11      0.00190 **
## Year2012      -0.24491      0.06730      -3.64      0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.254, Adjusted R-squared:  0.249
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 253 weights are ~= 1. The remaining 2871 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.202  0.869  0.944  0.901  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.20e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## LastAuthorFemale  1.026 1      1.013
## Year              1.040 16      1.001

```

## Residuals from first and last author



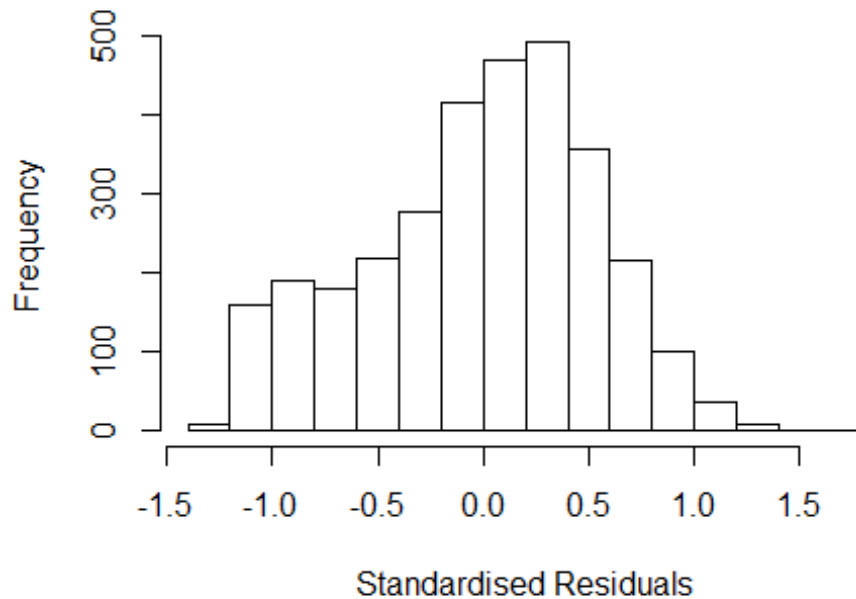
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2659 -0.3661 0.0493 0.3719 1.8396
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2001 0.0516 23.28 < 2e-16 ***
## FirstAuthorFemale1 0.0426 0.0210 2.03 0.04221 *
## LastAuthorFemale1 -0.0731 0.0267 -2.73 0.00629 **
## Year1997 0.0657 0.0707 0.93 0.35281
## Year1998 0.0982 0.0632 1.55 0.12053
## Year1999 -0.0596 0.0681 -0.88 0.38151
## Year2000 -0.0262 0.0648 -0.40 0.68593
## Year2001 -0.0569 0.0649 -0.88 0.38141
## Year2002 -0.0563 0.0682 -0.83 0.40928
## Year2003 -0.2407 0.0700 -3.44 0.00059 ***
## Year2004 -0.2932 0.0667 -4.40 1.1e-05 ***
## Year2005 -0.1390 0.0616 -2.26 0.02411 *
```

```

## Year2006          -0.2535      0.0648   -3.91  9.5e-05 ***
## Year2007          -0.1837      0.0629   -2.92  0.00351 **
## Year2008          -0.2737      0.0647   -4.23  2.4e-05 ***
## Year2009          -0.2112      0.0643   -3.28  0.00103 **
## Year2010          -0.2132      0.0670   -3.18  0.00148 **
## Year2011          -0.1732      0.0635   -2.73  0.00645 **
## Year2012          -0.2002      0.0661   -3.03  0.00246 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0399, Adjusted R-squared:  0.0343
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 237 weights are ~= 1. The remaining 2887 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.854  0.945  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.016 1      1.008
## Year              1.016 16      1.000

```

## Residuals from first author



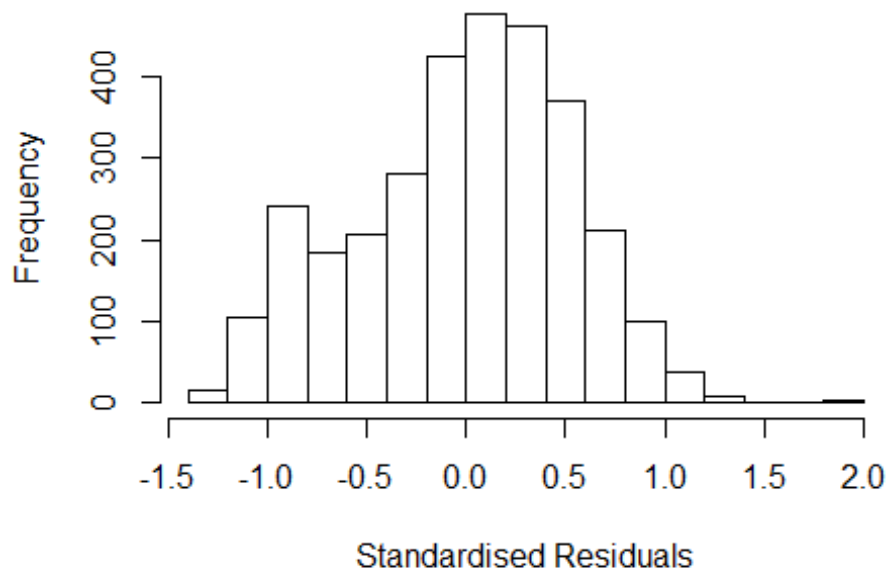
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2570 -0.3772 0.0489 0.3711 1.7908
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1923 0.0511 23.35 < 2e-16 ***
## FirstAuthorFemale1 0.0330 0.0211 1.57 0.11761
## Year1997 0.0647 0.0702 0.92 0.35738
## Year1998 0.0970 0.0628 1.55 0.12228
## Year1999 -0.0611 0.0677 -0.90 0.36676
## Year2000 -0.0276 0.0643 -0.43 0.66825
## Year2001 -0.0632 0.0646 -0.98 0.32793
## Year2002 -0.0597 0.0680 -0.88 0.37989
## Year2003 -0.2373 0.0694 -3.42 0.00064 ***
## Year2004 -0.2963 0.0666 -4.45 8.9e-06 ***
## Year2005 -0.1419 0.0613 -2.32 0.02064 *
## Year2006 -0.2550 0.0646 -3.95 8.1e-05 ***
```

```

## Year2007          -0.1859      0.0627   -2.97  0.00305 **
## Year2008          -0.2792      0.0644   -4.33  1.5e-05 ***
## Year2009          -0.2180      0.0641   -3.40  0.00068 ***
## Year2010          -0.2190      0.0670   -3.27  0.00109 **
## Year2011          -0.1801      0.0633   -2.85  0.00447 **
## Year2012          -0.2071      0.0659   -3.15  0.00168 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0371, Adjusted R-squared:  0.0318
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 253 weights are ~= 1. The remaining 2871 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.219  0.853  0.945  0.902  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.010
## Year      1.021 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2778 -0.3711 0.0515 0.3777 1.8545
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2106 0.0517 23.41 < 2e-16 ***
## LastAuthorFemale1 -0.0649 0.0266 -2.44 0.01465 *
## Year1997 0.0673 0.0710 0.95 0.34364
## Year1998 0.1010 0.0636 1.59 0.11220
## Year1999 -0.0574 0.0685 -0.84 0.40245
## Year2000 -0.0216 0.0648 -0.33 0.73917
## Year2001 -0.0522 0.0654 -0.80 0.42449
## Year2002 -0.0529 0.0686 -0.77 0.44098
## Year2003 -0.2379 0.0707 -3.37 0.00077 ***
## Year2004 -0.2932 0.0672 -4.36 1.3e-05 ***
## Year2005 -0.1350 0.0619 -2.18 0.02938 *
## Year2006 -0.2483 0.0650 -3.82 0.00013 ***
```

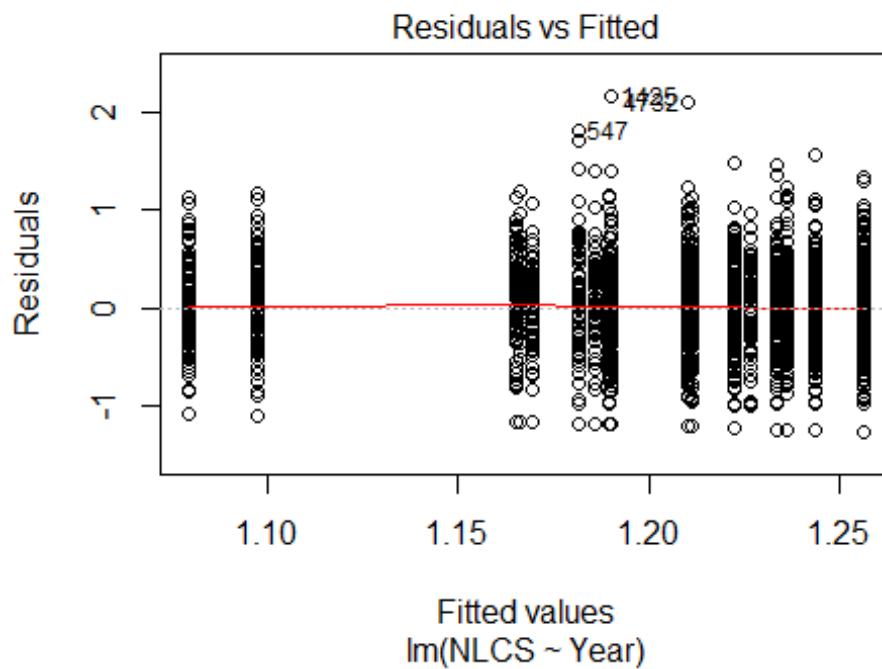
```

## Year2007          -0.1772      0.0632   -2.80   0.00511 **
## Year2008          -0.2672      0.0650   -4.11   4.0e-05 ***
## Year2009          -0.2056      0.0644   -3.19   0.00142 **
## Year2010          -0.2058      0.0672   -3.06   0.00223 **
## Year2011          -0.1665      0.0637   -2.61   0.00901 **
## Year2012          -0.1912      0.0662   -2.89   0.00392 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0386, Adjusted R-squared:  0.0334
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 238 weights are ~= 1. The remaining 2886 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.185  0.853  0.945  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3124"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2713"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 163 166 174 169 164 224 203 193 192 208 243 262 263 325 294
## 2011 2012
## 347 334
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 64 70 88 79 53 69 115 104 131 161 179 178 173 202 178
## 2011 2012

```



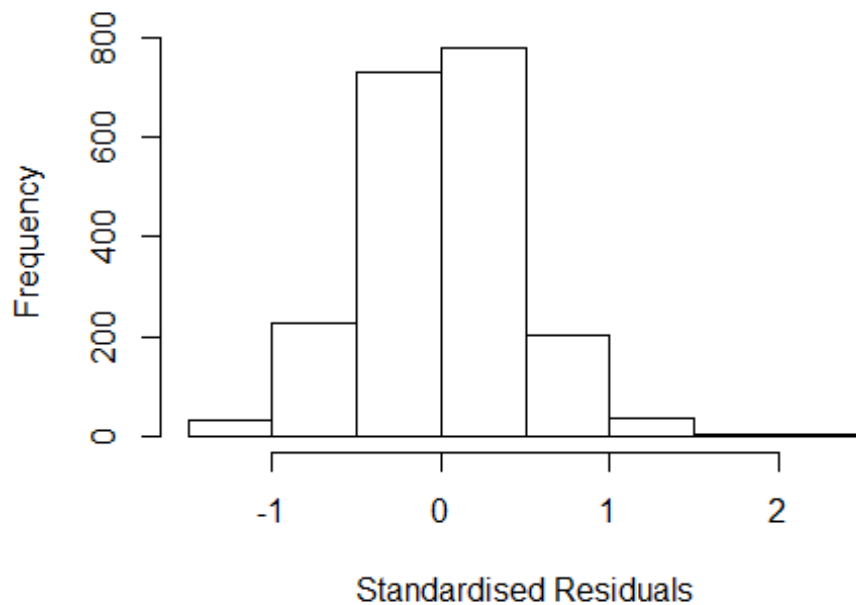
```
## 238 215
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 59 69 81 73 49 61 101 89 109 142 151 149 151 167 161
## 2011 2012
## 213 192
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 44, df = 16, p-value = 2e-04
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 21, df = 1, p-value = 5e-06
```



## Residuals from first and last author and team size



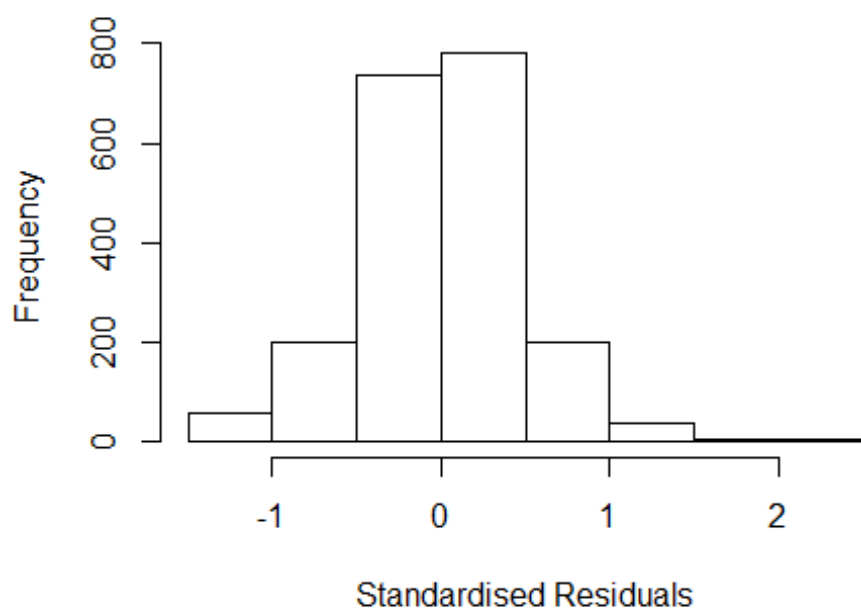
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33453 -0.26225 0.00722 0.27390 2.20086
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04409 0.06316 16.53 < 2e-16 ***
## FirstAuthorFemale1 0.01974 0.02091 0.94 0.3453
## LastAuthorFemale1 0.00689 0.02169 0.32 0.7509
## UniqueAuthors2 0.20573 0.04504 4.57 5.2e-06 ***
## UniqueAuthors3 0.25445 0.03917 6.50 1.0e-10 ***
## UniqueAuthors4 0.26435 0.04241 6.23 5.5e-10 ***
## UniqueAuthors5 0.30672 0.03704 8.28 2.2e-16 ***
## Year1997 -0.07025 0.07857 -0.89 0.3714
## Year1998 -0.12280 0.08638 -1.42 0.1553
## Year1999 -0.07656 0.08290 -0.92 0.3558
```

```

## Year2000      -0.03665      0.09267      -0.40      0.6926
## Year2001      -0.07092      0.07963      -0.89      0.3732
## Year2002      -0.09867      0.07131      -1.38      0.1666
## Year2003      -0.19939      0.07425      -2.69      0.0073 **
## Year2004      -0.10148      0.06965      -1.46      0.1453
## Year2005      -0.21060      0.07040      -2.99      0.0028 **
## Year2006      -0.02973      0.06731      -0.44      0.6587
## Year2007      -0.06745      0.06983      -0.97      0.3342
## Year2008      -0.05762      0.06899      -0.84      0.4037
## Year2009      -0.03637      0.06684      -0.54      0.5864
## Year2010      -0.07866      0.06820      -1.15      0.2489
## Year2011      -0.08833      0.06911      -1.28      0.2014
## Year2012      -0.04291      0.07009      -0.61      0.5405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.0711, Adjusted R-squared:  0.0608
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(152,431,1791) are outliers with |weight| = 0 ( < 5e-05);
## 182 weights are ~= 1. The remaining 1832 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0021 0.8490 0.9510 0.8890 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max      maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.129 1      1.062
## LastAuthorFemale 1.085 1      1.042
## Year      1.109 16      1.003

```

## Residuals from first and last author

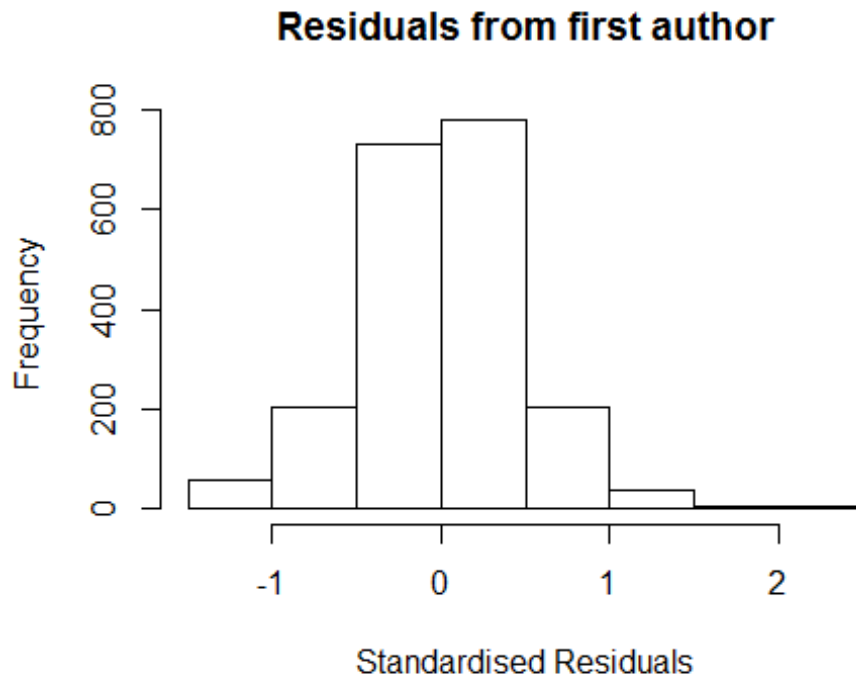


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30292 -0.27233 0.00944 0.27657 2.21871
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20e+00 5.59e-02 21.47 <2e-16 ***
## FirstAuthorFemale1 5.71e-02 2.07e-02 2.75 0.0059 **
## LastAuthorFemale1 1.84e-02 2.17e-02 0.85 0.3976
## Year1997 -2.42e-02 7.68e-02 -0.31 0.7531
## Year1998 -8.13e-02 8.65e-02 -0.94 0.3472
## Year1999 -3.52e-02 8.28e-02 -0.42 0.6711
## Year2000 1.17e-02 9.38e-02 0.12 0.9008
## Year2001 -2.86e-02 7.96e-02 -0.36 0.7193
## Year2002 -6.61e-02 7.20e-02 -0.92 0.3589
## Year2003 -1.57e-01 7.40e-02 -2.12 0.0344 *
## Year2004 -3.61e-02 6.78e-02 -0.53 0.5939
## Year2005 -1.52e-01 6.92e-02 -2.19 0.0286 *
```

```

## Year2006          2.70e-02   6.35e-02   0.43   0.6709
## Year2007          2.42e-05   6.72e-02   0.00   0.9997
## Year2008          2.25e-02   6.63e-02   0.34   0.7339
## Year2009          3.30e-02   6.46e-02   0.51   0.6097
## Year2010         -2.07e-03   6.46e-02  -0.03   0.9744
## Year2011         -1.88e-02   6.62e-02  -0.28   0.7765
## Year2012          2.80e-02   6.79e-02   0.41   0.6797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0225, Adjusted R-squared:  0.0137
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(431,1791) are outliers with |weight| = 0 ( < 5e-05);
## 165 weights are ~ = 1. The remaining 1850 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0169 0.8600 0.9520 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.081 1          1.040
## Year              1.081 16          1.002

```



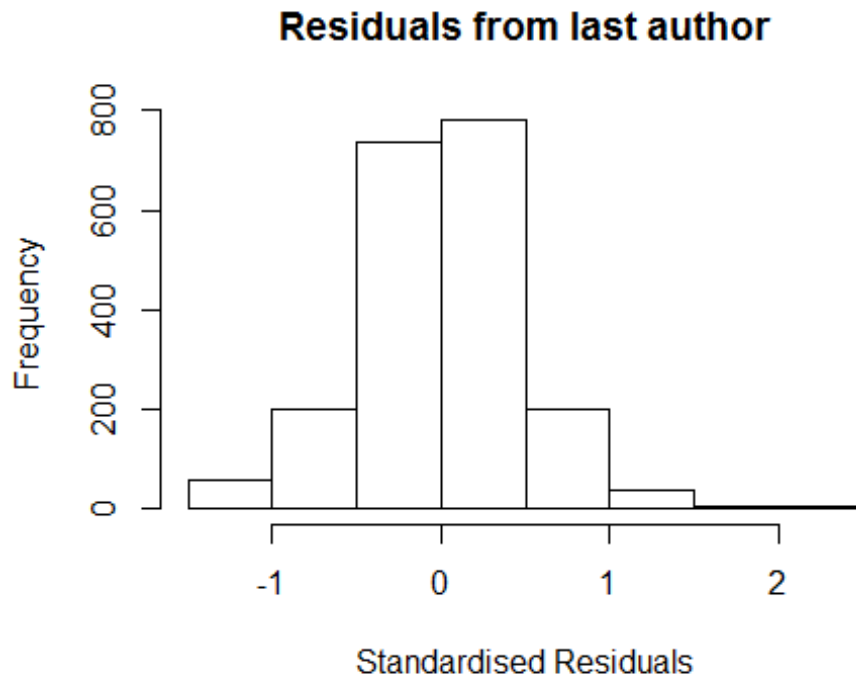
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29234 -0.26937 0.00824 0.27727 2.21566
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20263 0.05615 21.42 <2e-16 ***
## FirstAuthorFemale1 0.06104 0.02031 3.00 0.0027 **
## Year1997 -0.02559 0.07703 -0.33 0.7398
## Year1998 -0.08155 0.08667 -0.94 0.3468
## Year1999 -0.03393 0.08294 -0.41 0.6825
## Year2000 0.01128 0.09399 0.12 0.9045
## Year2001 -0.02887 0.07981 -0.36 0.7176
## Year2002 -0.06629 0.07228 -0.92 0.3592
## Year2003 -0.15637 0.07429 -2.10 0.0354 *
## Year2004 -0.03489 0.06804 -0.51 0.6081
## Year2005 -0.15068 0.06954 -2.17 0.0304 *
## Year2006 0.02774 0.06379 0.43 0.6637
```

```

## Year2007          0.00238    0.06742    0.04    0.9718
## Year2008          0.02339    0.06659    0.35    0.7254
## Year2009          0.03459    0.06483    0.53    0.5937
## Year2010         -0.00157    0.06485   -0.02    0.9807
## Year2011         -0.01670    0.06639   -0.25    0.8014
## Year2012          0.02868    0.06820    0.42    0.6741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.0139
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(431,1791) are outliers with |weight| = 0 ( < 5e-05);
## 164 weights are ~= 1. The remaining 1851 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.019  0.860   0.952   0.891   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.039  1         1.020
## Year             1.039 16         1.001

```





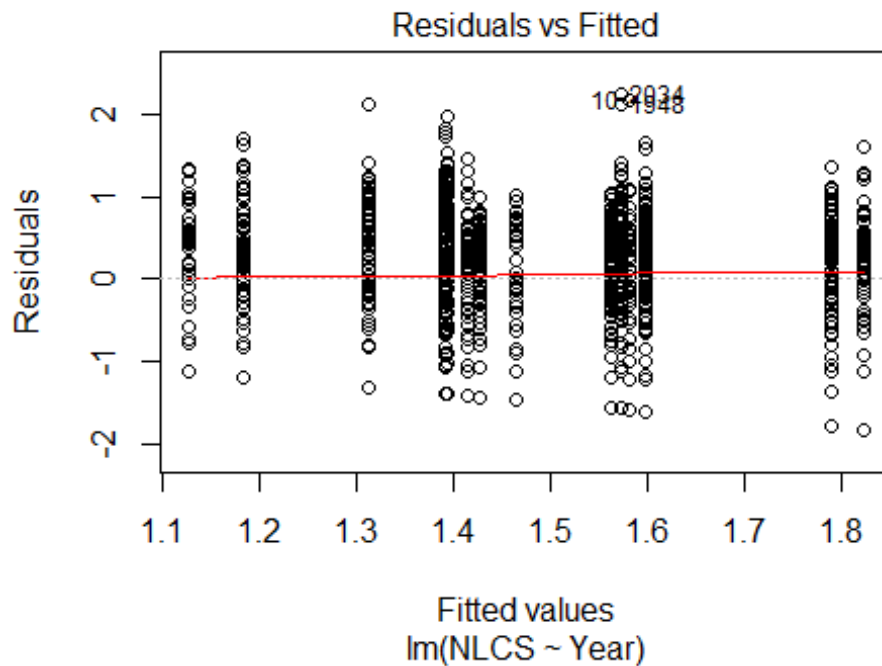
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2830 -0.2719 0.0131 0.2768 2.2045
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21279 0.05715 21.22 <2e-16 ***
## LastAuthorFemale1 0.03209 0.02128 1.51 0.132
## Year1997 -0.01468 0.07768 -0.19 0.850
## Year1998 -0.07743 0.08686 -0.89 0.373
## Year1999 -0.03074 0.08335 -0.37 0.712
## Year2000 0.01600 0.09586 0.17 0.867
## Year2001 -0.01684 0.08049 -0.21 0.834
## Year2002 -0.06526 0.07304 -0.89 0.372
## Year2003 -0.15047 0.07486 -2.01 0.045 *
## Year2004 -0.03183 0.06859 -0.46 0.643
## Year2005 -0.14548 0.07004 -2.08 0.038 *
## Year2006 0.03283 0.06447 0.51 0.611
```

```

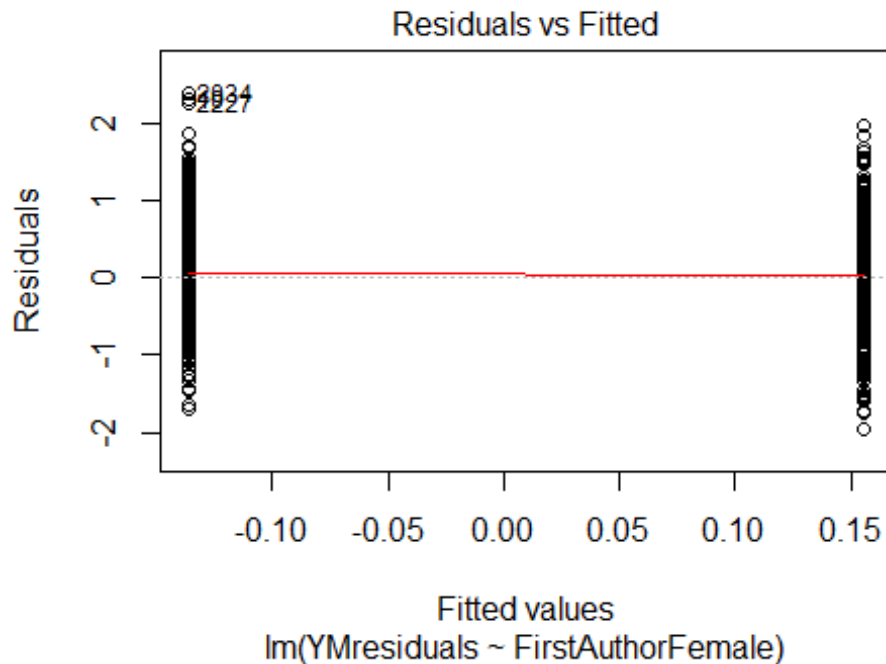
## Year2007          0.00529      0.06783      0.08      0.938
## Year2008          0.03238      0.06684      0.48      0.628
## Year2009          0.04309      0.06532      0.66      0.510
## Year2010          0.01223      0.06514      0.19      0.851
## Year2011         -0.00762      0.06681     -0.11      0.909
## Year2012          0.03814      0.06842      0.56      0.577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.417
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.0102
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(431,1791) are outliers with |weight| = 0 ( < 5e-05);
## 165 weights are ~= 1. The remaining 1850 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0092 0.8580 0.9520 0.8910 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.96e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2017"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2714"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   90   64  113   76  104  123  109  104  120   93  114  130   93  130  108
## 2011 2012
##  140  159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   56   40   94   53   39   36   87   84  101   78   96  110   78  120   98

```

```
## 2011 2012
## 125 138
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 35 87 51 36 33 71 77 92 75 89 98 73 114 92
## 2011 2012
## 118 130
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 67, df = 16, p-value = 3e-08
```

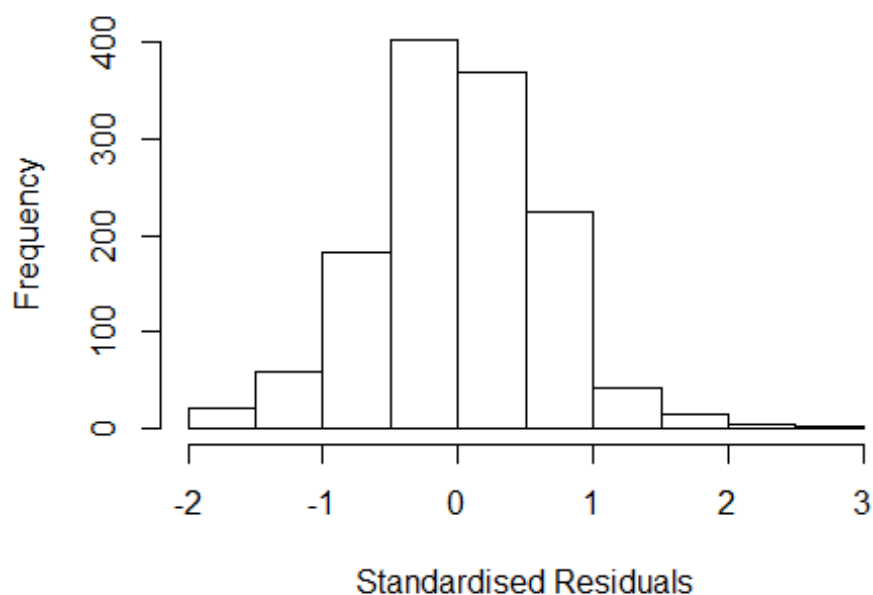


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 20, df = 1, p-value = 9e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.00714009353141"
## [1] "Male first author team size 2018 geometric mean: 3.55394056962602"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 980, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.95436346521196"
## [1] "Male last author team size 2018 geometric mean: 3.7837353161035"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.156 1      1.075
## LastAuthorFemale  1.156 1      1.075
## UniqueAuthors    1.716 4      1.070
## Year              1.767 16     1.018
```

## Residuals from first and last author and team size



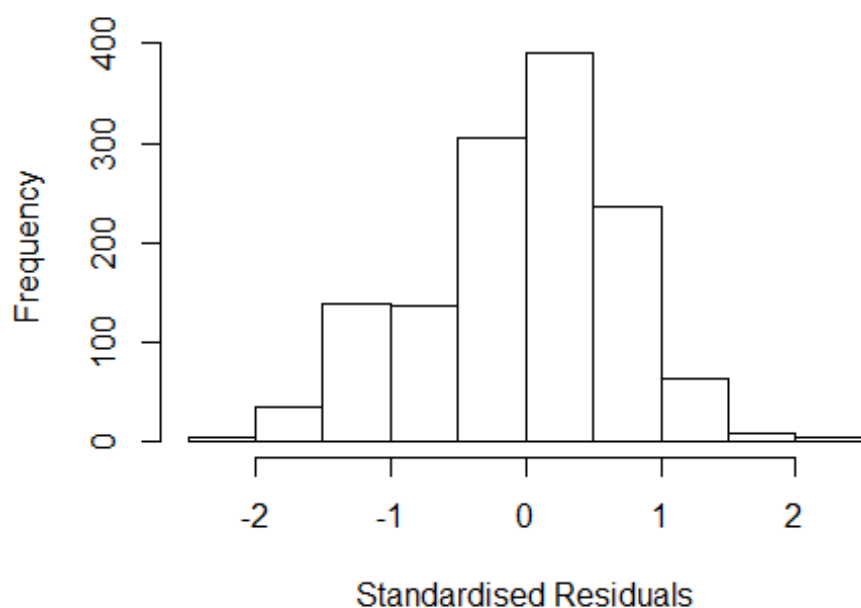
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 10 0030453840 3.754 1996      2714      1      2.938
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.91434 -0.42239 -0.00241  0.40306  2.93787
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.81613    0.14821   5.51 4.4e-08 ***
## FirstAuthorFemale1 0.17357    0.03719   4.67 3.4e-06 ***
## LastAuthorFemale1 0.09677    0.03916   2.47  0.0136 *
## UniqueAuthors2    0.81925    0.08602   9.52 < 2e-16 ***
## UniqueAuthors3    0.97139    0.07365  13.19 < 2e-16 ***
## UniqueAuthors4    1.00259    0.07100  14.12 < 2e-16 ***
## UniqueAuthors5    1.11090    0.06520  17.04 < 2e-16 ***
## Year1997          0.05667    0.23409   0.24  0.8087
## Year1998         -0.38270    0.14914  -2.57  0.0104 *
## Year1999         -0.44681    0.16175  -2.76  0.0058 **
```

```

## Year2000      -0.23707      0.16212      -1.46      0.1439
## Year2001      -0.35492      0.15983      -2.22      0.0266 *
## Year2002      -0.38948      0.14571      -2.67      0.0076 **
## Year2003      -0.44268      0.14909      -2.97      0.0030 **
## Year2004      -0.41387      0.14586      -2.84      0.0046 **
## Year2005      -0.28884      0.14684      -1.97      0.0494 *
## Year2006      -0.17272      0.14630      -1.18      0.2380
## Year2007       0.00861      0.15370       0.06      0.9553
## Year2008      -0.00224      0.15358      -0.01      0.9883
## Year2009      -0.32342      0.14964      -2.16      0.0309 *
## Year2010      -0.24178      0.15924      -1.52      0.1292
## Year2011      -0.18975      0.14666      -1.29      0.1960
## Year2012      -0.36336      0.14569      -2.49      0.0128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.574
## Multiple R-squared:  0.362, Adjusted R-squared:  0.352
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## observation 4 is an outlier with |weight| = 0 ( < 7.6e-05);
## 102 weights are ~= 1. The remaining 1217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0727 0.8600 0.9440 0.8880 0.9830 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           7.58e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.094 1 1.046
## LastAuthorFemale 1.073 1 1.036
## Year 1.174 16 1.005

```

## Residuals from first and last author



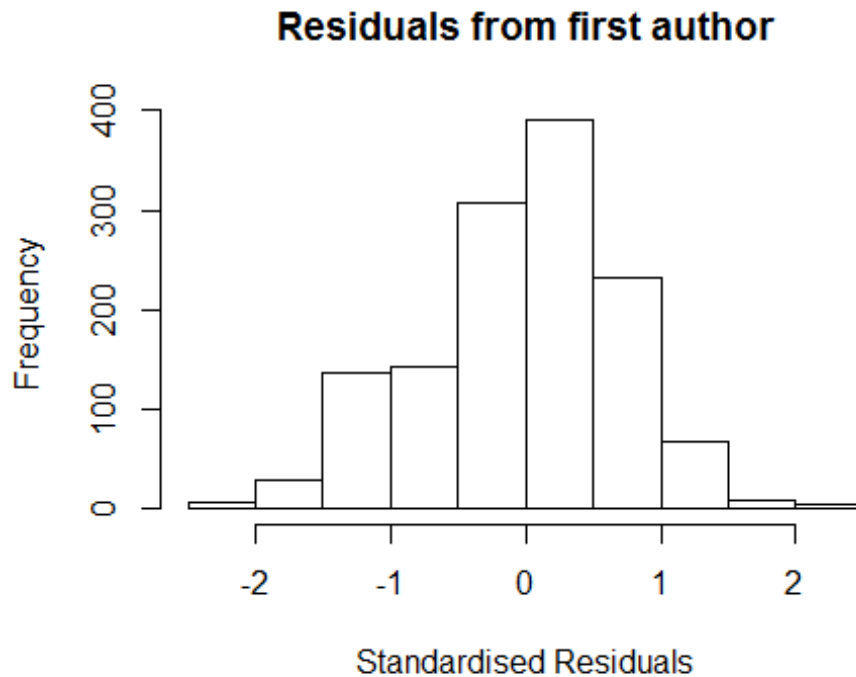
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.0712 -0.4511 0.0596 0.4732 2.3481
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4998 0.0931 16.11 < 2e-16 ***
## FirstAuthorFemale1 0.2843 0.0420 6.77 1.9e-11 ***
## LastAuthorFemale1 0.0882 0.0424 2.08 0.03798 *
## Year1997 0.0724 0.2007 0.36 0.71824
## Year1998 -0.4372 0.1254 -3.49 0.00051 ***
## Year1999 -0.5319 0.1465 -3.63 0.00029 ***
## Year2000 -0.2390 0.1366 -1.75 0.08038 .
## Year2001 -0.1161 0.1291 -0.90 0.36861
## Year2002 -0.1901 0.1100 -1.73 0.08409 .
## Year2003 -0.2501 0.1153 -2.17 0.03031 *
## Year2004 -0.2634 0.1084 -2.43 0.01526 *
## Year2005 -0.0634 0.1173 -0.54 0.58871
```

```

## Year2006          -0.0389      0.1156   -0.34  0.73638
## Year2007           0.1990      0.1230    1.62  0.10591
## Year2008           0.2184      0.1185    1.84  0.06559 .
## Year2009          -0.2816      0.1282   -2.20  0.02824 *
## Year2010          -0.2238      0.1389   -1.61  0.10752
## Year2011          -0.0319      0.1198   -0.27  0.78995
## Year2012          -0.3421      0.1190   -2.88  0.00409 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.686
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.0974
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 116 weights are ~= 1. The remaining 1204 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.217  0.846  0.949  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.102 1      1.050
## Year              1.102 16      1.003

```



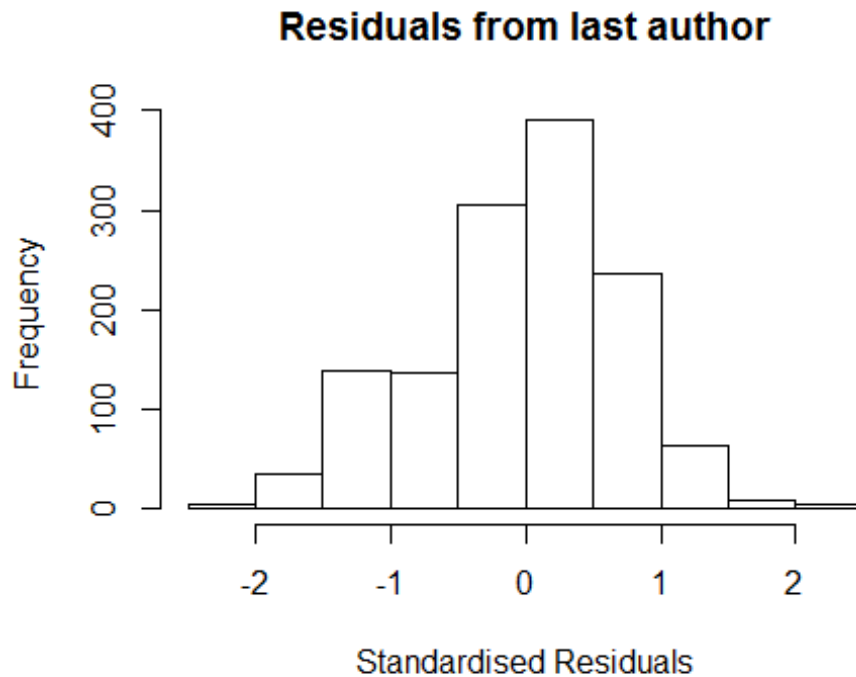


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0295 -0.4481  0.0504  0.4778  2.3309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5283     0.0904   16.91 < 2e-16 ***
## FirstAuthorFemale1  0.3014     0.0433    6.96 5.2e-12 ***
## Year1997          0.0951     0.1982    0.48 0.63141
## Year1998         -0.4422     0.1251   -3.53 0.00042 ***
## Year1999         -0.5362     0.1475   -3.63 0.00029 ***
## Year2000         -0.2396     0.1380   -1.74 0.08264 .
## Year2001         -0.1012     0.1271   -0.80 0.42611
## Year2002         -0.1959     0.1089   -1.80 0.07230 .
## Year2003         -0.2670     0.1143   -2.33 0.01970 *
## Year2004         -0.2681     0.1078   -2.49 0.01297 *
## Year2005         -0.0656     0.1168   -0.56 0.57472
## Year2006         -0.0389     0.1148   -0.34 0.73445
```

```

## Year2007          0.1998      0.1217      1.64  0.10082
## Year2008          0.2096      0.1173      1.79  0.07429 .
## Year2009         -0.2962      0.1285     -2.31  0.02129 *
## Year2010         -0.2321      0.1391     -1.67  0.09549 .
## Year2011         -0.0432      0.1189     -0.36  0.71670
## Year2012         -0.3493      0.1183     -2.95  0.00320 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.672
## Multiple R-squared:  0.108, Adjusted R-squared:  0.0959
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.205  0.842  0.947  0.891  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.071 1          1.035
## Year            1.071 16          1.002

```



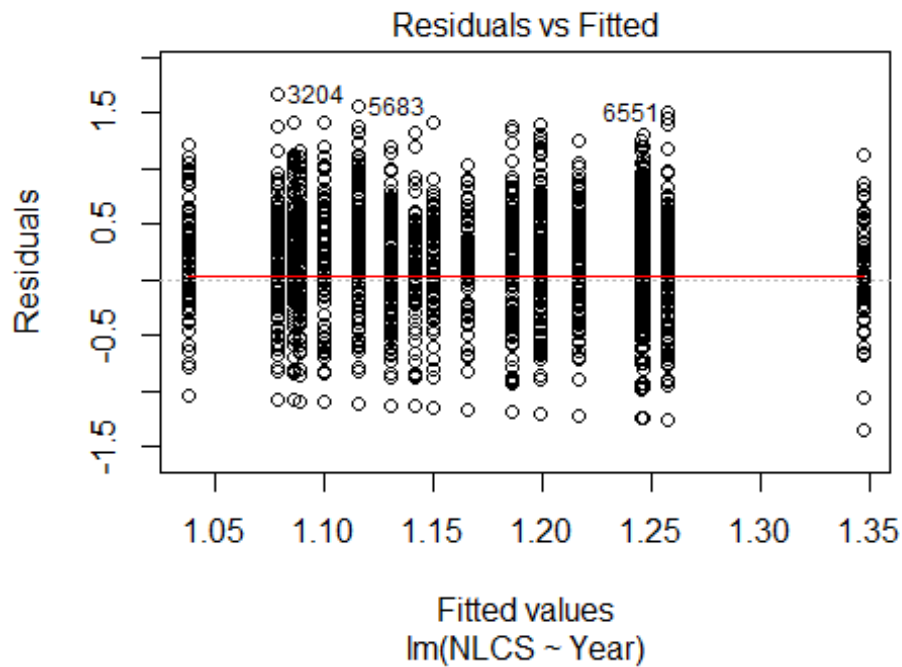
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9940 -0.4534  0.0741  0.4790  2.2216
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.58525    0.09226   17.18 < 2e-16 ***
## LastAuthorFemale1 0.14992    0.04337    3.46 0.00056 ***
## Year1997         0.09012    0.19762    0.46 0.64845
## Year1998        -0.45012    0.12725   -3.54 0.00042 ***
## Year1999        -0.51792    0.15289   -3.39 0.00073 ***
## Year2000        -0.17464    0.13834   -1.26 0.20702
## Year2001        -0.05970    0.12920   -0.46 0.64411
## Year2002        -0.17143    0.10902   -1.57 0.11608
## Year2003        -0.20962    0.11788   -1.78 0.07560 .
## Year2004        -0.23507    0.10852   -2.17 0.03048 *
## Year2005        -0.01534    0.11628   -0.13 0.89503
## Year2006         0.00533    0.11661    0.05 0.96358
```

```

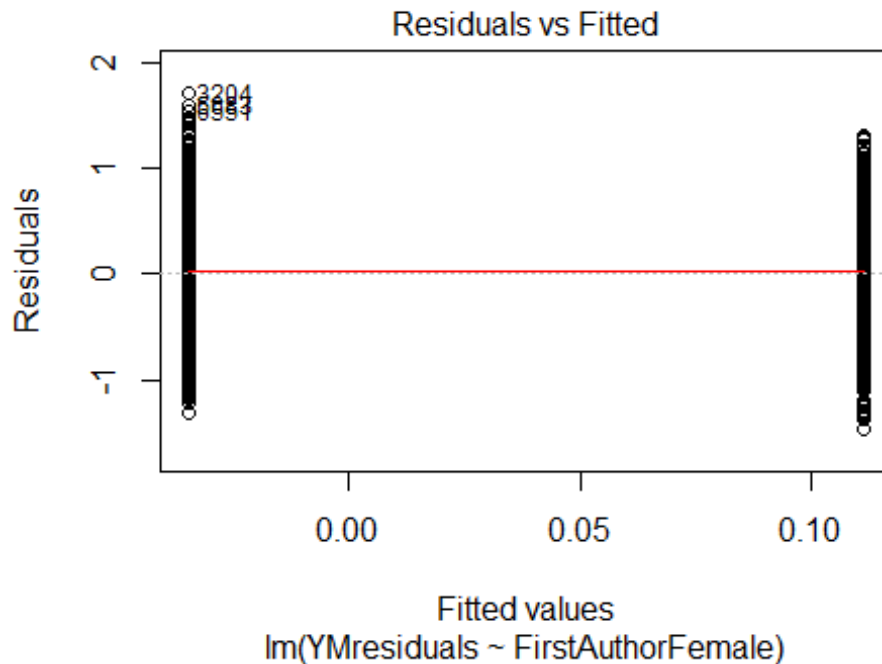
## Year2007      0.21738      0.12241      1.78  0.07600 .
## Year2008      0.25880      0.11961      2.16  0.03068 *
## Year2009     -0.21426      0.13365     -1.60  0.10913
## Year2010     -0.18542      0.14591     -1.27  0.20404
## Year2011      0.00913      0.12181      0.07  0.94024
## Year2012     -0.28023      0.12095     -2.32  0.02066 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.675
## Multiple R-squared:  0.0786, Adjusted R-squared:  0.0666
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 1225 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.256  0.847  0.947  0.889  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1320"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2715"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 331 318 305 320 375 310 331 278 283 321 326 377 366 336 308
## 2011 2012
## 395 371
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 84 88 95 109 127 135 134 128 158 162 203 179 177 166
## 2011 2012

```

```
## 195 184
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 55 69 83 77 91 98 111 100 98 126 134 153 141 131 141
## 2011 2012
## 162 158
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.009
```

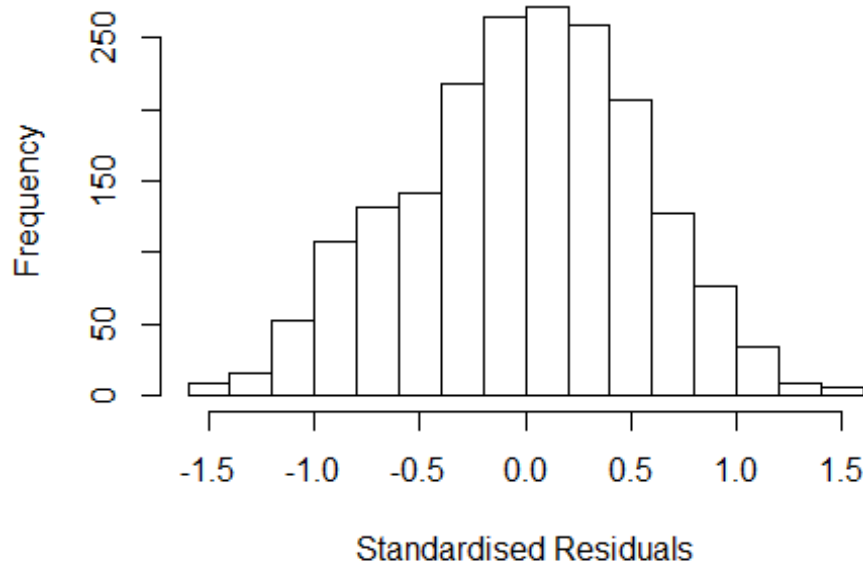


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 5.15930348364289"
## [1] "Male first author team size 2018 geometric mean: 5.56620589380313"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.6107910497769"
## [1] "Male last author team size 2018 geometric mean: 5.40058688772426"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1          1.032
## LastAuthorFemale  1.053 1          1.026
## UniqueAuthors    1.231 4          1.026
## Year              1.309 16         1.008
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5241 -0.3594  0.0231  0.3843  1.5838
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.99075    0.08342   11.88 < 2e-16 ***
## FirstAuthorFemale1 0.09534    0.02972    3.21 0.00136 **
## LastAuthorFemale1 0.00405    0.04273    0.09 0.92459
## UniqueAuthors2    0.05541    0.06651    0.83 0.40494
## UniqueAuthors3    0.27180    0.06185    4.39 1.2e-05 ***
## UniqueAuthors4    0.31803    0.05797    5.49 4.7e-08 ***
## UniqueAuthors5    0.56139    0.05303   10.59 < 2e-16 ***
## Year1997         -0.11232    0.09438   -1.19 0.23415
## Year1998         -0.13166    0.09190   -1.43 0.15212
## Year1999         -0.12001    0.08945   -1.34 0.17987
```

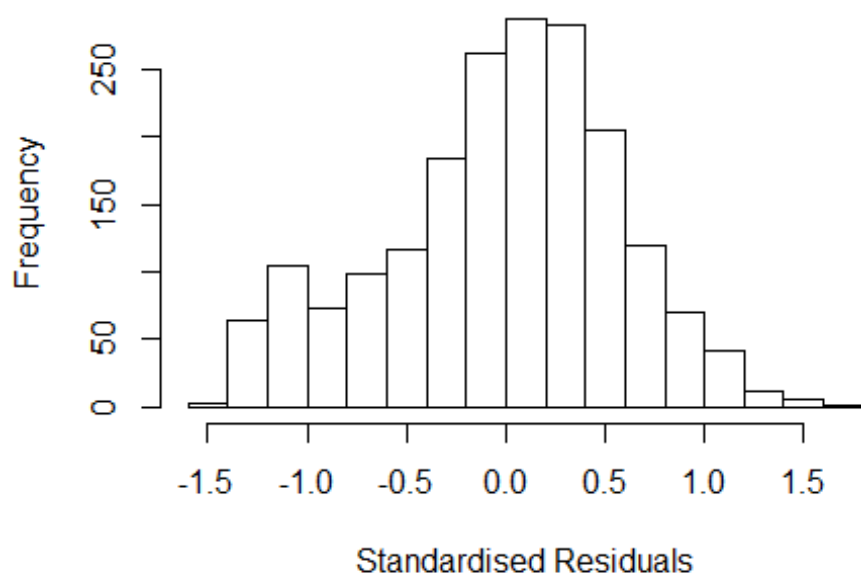
```

## Year2000      -0.23762    0.09200    -2.58    0.00988 **
## Year2001      -0.21733    0.09577    -2.27    0.02336 *
## Year2002      -0.26583    0.08975    -2.96    0.00309 **
## Year2003      -0.31555    0.08871    -3.56    0.00038 ***
## Year2004      -0.25733    0.09206    -2.80    0.00524 **
## Year2005      -0.22219    0.08101    -2.74    0.00615 **
## Year2006      -0.17559    0.08201    -2.14    0.03239 *
## Year2007      -0.12339    0.08567    -1.44    0.14995
## Year2008      -0.12851    0.08416    -1.53    0.12695
## Year2009      -0.13997    0.08500    -1.65    0.09976 .
## Year2010      -0.22614    0.08693    -2.60    0.00935 **
## Year2011      -0.14149    0.08387    -1.69    0.09177 .
## Year2012      -0.07494    0.08132    -0.92    0.35688
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.151, Adjusted R-squared:  0.142
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1787 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.372  0.852  0.949   0.905   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1      1.029
## LastAuthorFemale  1.065 1      1.032
## Year              1.119 16      1.004

```



## Residuals from first and last author



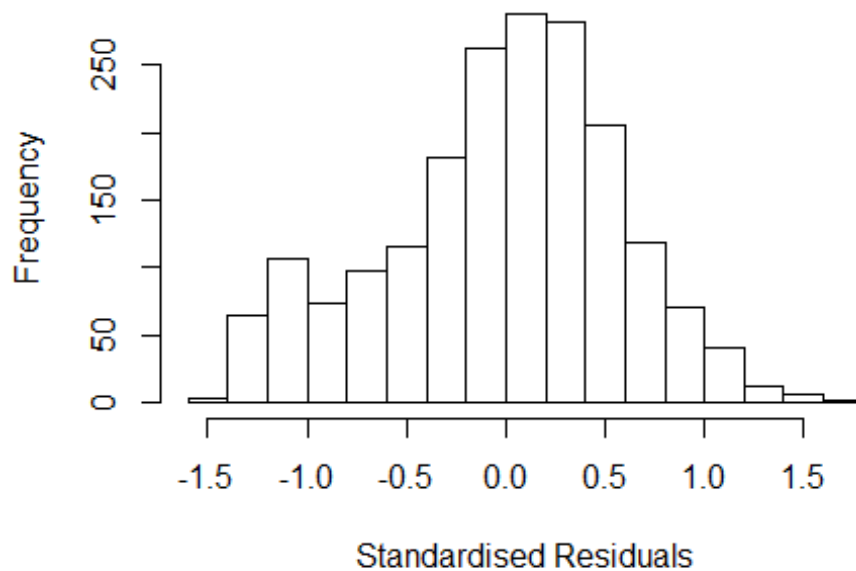
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4163 -0.3611 0.0376 0.3693 1.6402
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.32689 0.05873 22.59 < 2e-16 ***
## FirstAuthorFemale1 0.14000 0.03096 4.52 6.5e-06 ***
## LastAuthorFemale1 -0.00198 0.04715 -0.04 0.96655
## Year1997 -0.18642 0.08414 -2.22 0.02684 *
## Year1998 -0.19355 0.08485 -2.28 0.02266 *
## Year1999 -0.14464 0.08003 -1.81 0.07085 .
## Year2000 -0.33496 0.09342 -3.59 0.00034 ***
## Year2001 -0.21505 0.09389 -2.29 0.02211 *
## Year2002 -0.27378 0.08956 -3.06 0.00227 **
## Year2003 -0.29880 0.08921 -3.35 0.00083 ***
## Year2004 -0.22014 0.09416 -2.34 0.01949 *
## Year2005 -0.20006 0.07577 -2.64 0.00835 **
```

```

## Year2006          -0.16865      0.07548      -2.23  0.02558 *
## Year2007          -0.08401      0.07771      -1.08  0.27981
## Year2008          -0.09266      0.07808      -1.19  0.23550
## Year2009          -0.10538      0.07624      -1.38  0.16710
## Year2010          -0.19456      0.08251      -2.36  0.01847 *
## Year2011          -0.06856      0.07803      -0.88  0.37971
## Year2012          -0.05059      0.07309      -0.69  0.48890
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0319, Adjusted R-squared:  0.0228
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 159 weights are ~= 1. The remaining 1769 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.351  0.851  0.949  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07          5.19e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1          1.029
## Year              1.058 16          1.002

```

## Residuals from first author



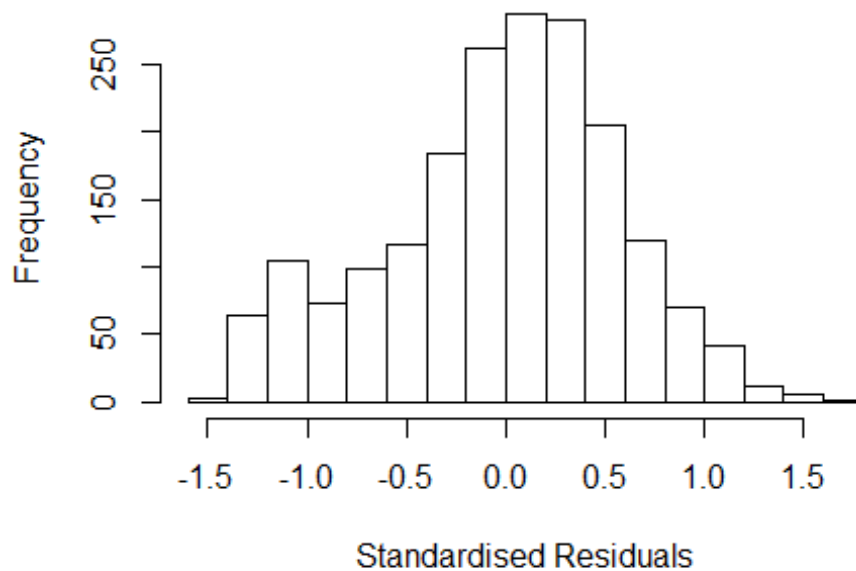
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4159 -0.3609 0.0377 0.3694 1.6403
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3268 0.0587 22.62 < 2e-16 ***
## FirstAuthorFemale1 0.1399 0.0310 4.51 6.9e-06 ***
## Year1997 -0.1864 0.0841 -2.22 0.02685 *
## Year1998 -0.1937 0.0846 -2.29 0.02220 *
## Year1999 -0.1447 0.0800 -1.81 0.07064 .
## Year2000 -0.3349 0.0934 -3.58 0.00035 ***
## Year2001 -0.2150 0.0939 -2.29 0.02216 *
## Year2002 -0.2739 0.0892 -3.07 0.00217 **
## Year2003 -0.2988 0.0892 -3.35 0.00083 ***
## Year2004 -0.2201 0.0941 -2.34 0.01946 *
## Year2005 -0.2001 0.0758 -2.64 0.00838 **
## Year2006 -0.1686 0.0755 -2.23 0.02561 *
```

```

## Year2007          -0.0840      0.0777   -1.08  0.28006
## Year2008          -0.0928      0.0780   -1.19  0.23461
## Year2009          -0.1054      0.0763   -1.38  0.16708
## Year2010          -0.1947      0.0825   -2.36  0.01835 *
## Year2011          -0.0685      0.0780   -0.88  0.38008
## Year2012          -0.0507      0.0731   -0.69  0.48757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.548
## Multiple R-squared:  0.0319, Adjusted R-squared:  0.0233
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 159 weights are ~= 1. The remaining 1769 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.350  0.851  0.949  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.062 1          1.031
## Year            1.062 16          1.002

```

## Residuals from last author



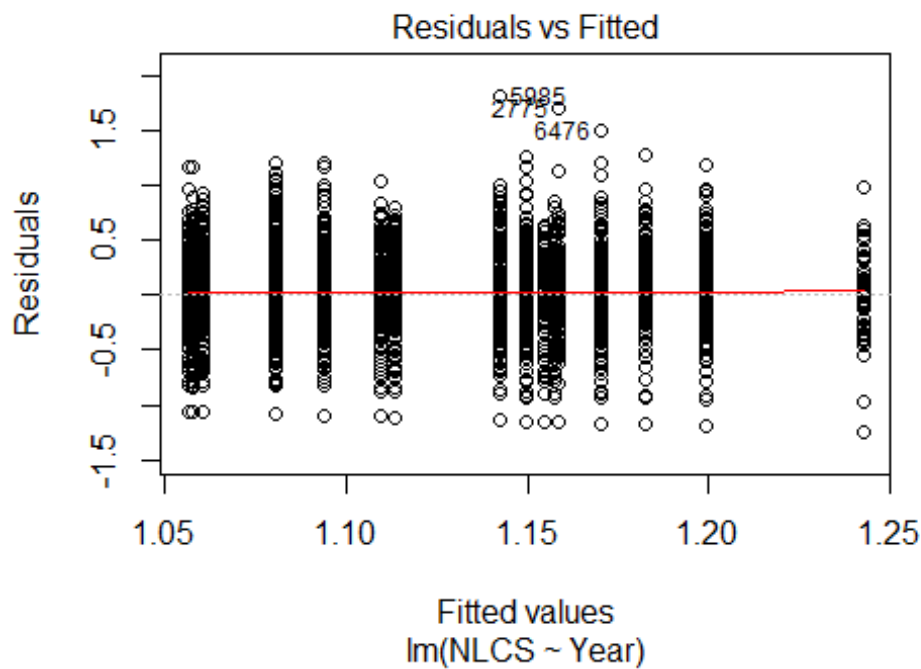
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3575 -0.3683 0.0448 0.3754 1.6099
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3575 0.0581 23.37 < 2e-16 ***
## LastAuthorFemale1 0.0207 0.0465 0.45 0.65598
## Year1997 -0.1902 0.0844 -2.26 0.02424 *
## Year1998 -0.1826 0.0834 -2.19 0.02866 *
## Year1999 -0.1575 0.0795 -1.98 0.04784 *
## Year2000 -0.3326 0.0951 -3.50 0.00048 ***
## Year2001 -0.2095 0.0940 -2.23 0.02593 *
## Year2002 -0.2795 0.0902 -3.10 0.00197 **
## Year2003 -0.2917 0.0902 -3.24 0.00124 **
## Year2004 -0.2204 0.0948 -2.33 0.02014 *
## Year2005 -0.2057 0.0758 -2.71 0.00675 **
## Year2006 -0.1645 0.0750 -2.19 0.02834 *
```

```

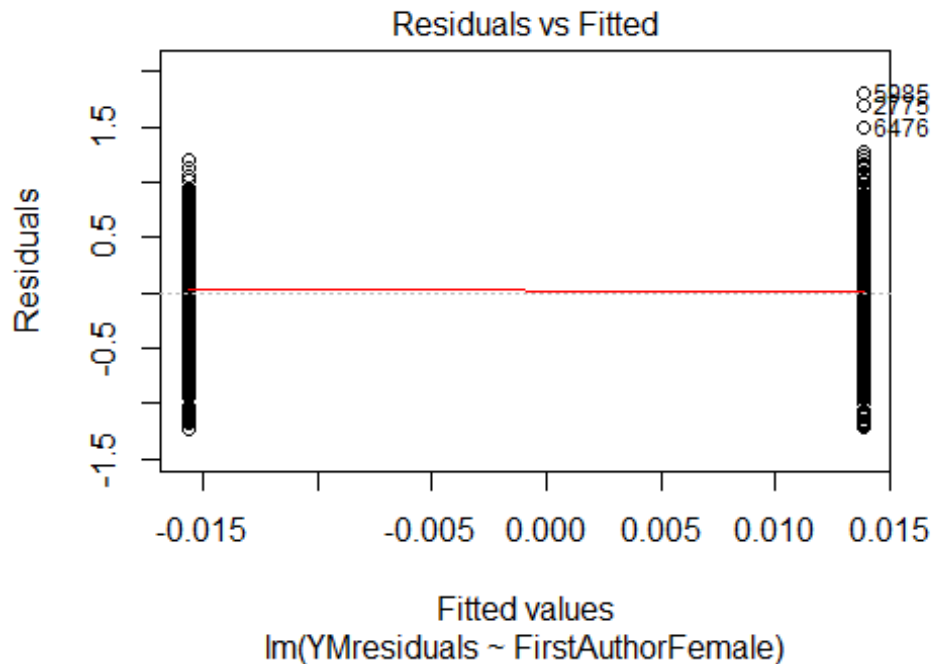
## Year2007          -0.0816      0.0780   -1.05   0.29537
## Year2008          -0.0876      0.0778   -1.12   0.26076
## Year2009          -0.1016      0.0757   -1.34   0.17956
## Year2010          -0.1915      0.0837   -2.29   0.02218 *
## Year2011          -0.0603      0.0783   -0.77   0.44120
## Year2012          -0.0428      0.0729   -0.59   0.55743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.549
## Multiple R-squared:  0.0214, Adjusted R-squared:  0.0127
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 159 weights are ~= 1. The remaining 1769 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.370  0.850  0.947  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.19e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1928"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2716"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  338  382  425  274  489  361  298  271  295  314  317  334  301  293  296
## 2011 2012
##  357  308
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  134  164   46  141   71  108  159  145  179  194  213  199  198  185  195
## 2011 2012

```

```
## 248 219
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 147 43 121 60 92 130 121 155 159 182 172 169 166 169
## 2011 2012
## 217 188
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 34, df = 16, p-value = 0.005
```



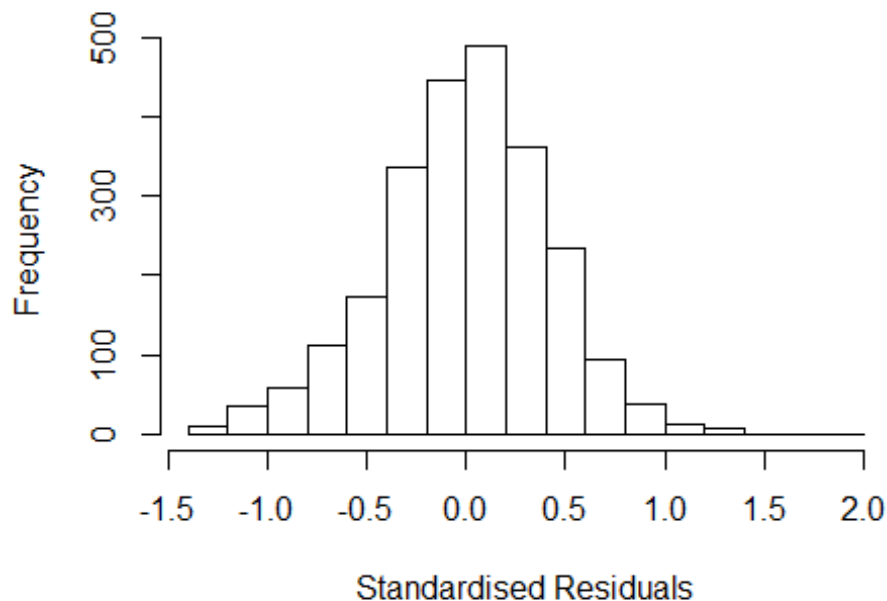
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 9e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 5.28504406443117"
## [1] "Male first author team size 2018 geometric mean: 3.79516715827415"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4500, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.18099562398562"
## [1] "Male last author team size 2018 geometric mean: 4.20284307876252"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4200, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.068 1 1.034
## LastAuthorFemale 1.046 1 1.023
## UniqueAuthors 1.222 4 1.025
## Year 1.260 16 1.007
```



## Residuals from first and last author and team size



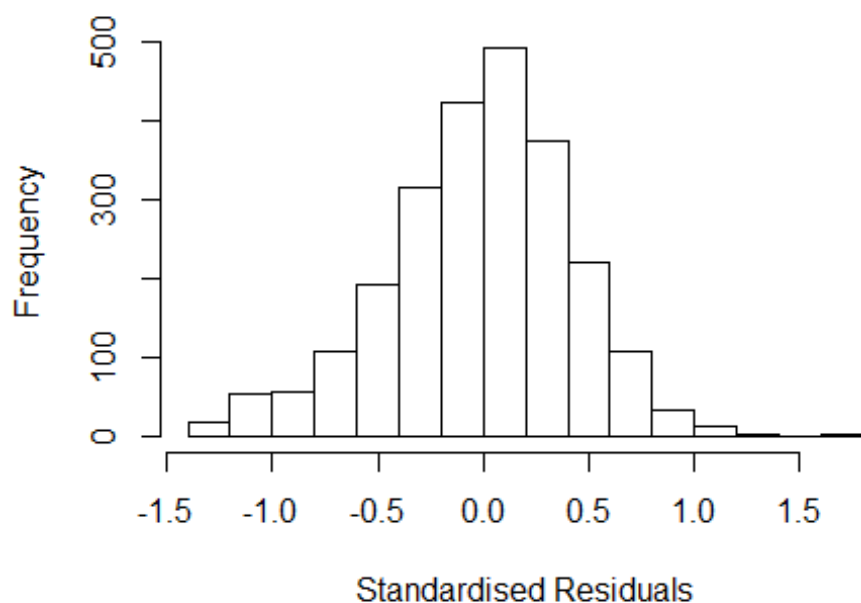
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3145 -0.2673 0.0146 0.2631 1.8519
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0589 0.0560 18.90 < 2e-16 ***
## FirstAuthorFemale1 -0.0422 0.0177 -2.39 0.01703 *
## LastAuthorFemale1 -0.0623 0.0193 -3.22 0.00131 **
## UniqueAuthors2 0.1482 0.0448 3.31 0.00096 ***
## UniqueAuthors3 0.2031 0.0429 4.73 2.3e-06 ***
## UniqueAuthors4 0.1858 0.0426 4.36 1.3e-05 ***
## UniqueAuthors5 0.3076 0.0374 8.23 3.0e-16 ***
## Year1997 -0.0152 0.0593 -0.26 0.79781
## Year1998 0.0993 0.0801 1.24 0.21535
## Year1999 -0.1086 0.0562 -1.93 0.05338 .
```

```

## Year2000      -0.0263      0.0701      -0.38      0.70757
## Year2001      -0.0980      0.0669      -1.47      0.14292
## Year2002      -0.0519      0.0577      -0.90      0.36887
## Year2003      -0.1714      0.0573      -2.99      0.00278 **
## Year2004      -0.1249      0.0540      -2.31      0.02086 *
## Year2005      -0.1676      0.0546      -3.07      0.00215 **
## Year2006      -0.1154      0.0544      -2.12      0.03396 *
## Year2007      -0.1858      0.0544      -3.41      0.00065 ***
## Year2008      -0.1019      0.0562      -1.81      0.07014 .
## Year2009      -0.0878      0.0552      -1.59      0.11173
## Year2010      -0.1997      0.0559      -3.57      0.00036 ***
## Year2011      -0.1518      0.0535      -2.84      0.00458 **
## Year2012      -0.1604      0.0550      -2.92      0.00357 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.0741, Adjusted R-squared:  0.0655
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 199 weights are ~= 1. The remaining 2211 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8640 0.9490 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.15e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1 1.021
## LastAuthorFemale 1.036 1 1.018
## Year 1.070 16 1.002

```

## Residuals from first and last author



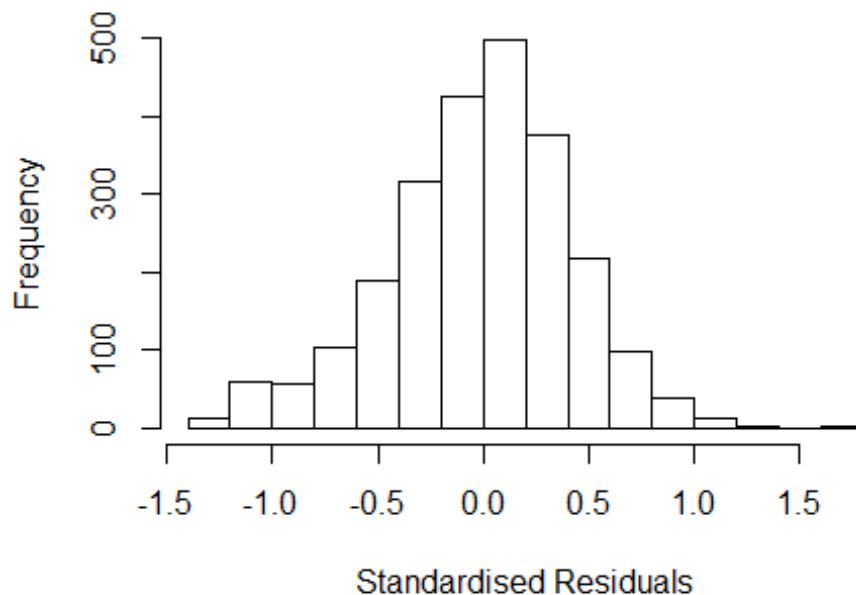
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2544 -0.2764  0.0152  0.2711  1.7816
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2423     0.0486   25.57 < 2e-16 ***
## FirstAuthorFemale1 -0.0161     0.0179   -0.90  0.36848
## LastAuthorFemale1  -0.0749     0.0197   -3.80  0.00015 ***
## Year1997           0.0121     0.0608    0.20  0.84190
## Year1998           0.0470     0.0773    0.61  0.54312
## Year1999          -0.0745     0.0586   -1.27  0.20419
## Year2000          -0.0328     0.0732   -0.45  0.65414
## Year2001          -0.0395     0.0688   -0.57  0.56582
## Year2002          -0.0291     0.0604   -0.48  0.63009
## Year2003          -0.1488     0.0604   -2.46  0.01378 *
## Year2004          -0.0955     0.0559   -1.71  0.08787 .
## Year2005          -0.1367     0.0564   -2.42  0.01539 *
```

```

## Year2006          -0.0699      0.0564   -1.24  0.21541
## Year2007          -0.1416      0.0563   -2.52  0.01193 *
## Year2008          -0.0668      0.0591   -1.13  0.25784
## Year2009          -0.0351      0.0569   -0.62  0.53709
## Year2010          -0.1528      0.0592   -2.58  0.00996 **
## Year2011          -0.1107      0.0562   -1.97  0.04911 *
## Year2012          -0.1123      0.0577   -1.95  0.05160 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.023, Adjusted R-squared:  0.0157
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 203 weights are ~= 1. The remaining 2207 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0159 0.8630 0.9510 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.15e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.04 1      1.020
## Year      1.04 16      1.001

```

## Residuals from first author



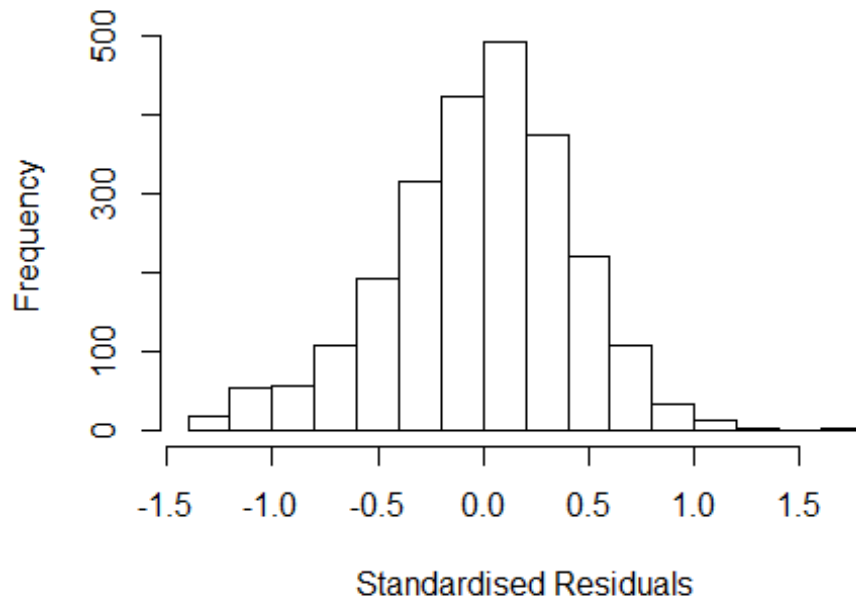
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2595 -0.2788 0.0176 0.2750 1.7982
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2264 0.0482 25.47 <2e-16 ***
## FirstAuthorFemale1 -0.0274 0.0180 -1.52 0.1290
## Year1997 0.0121 0.0609 0.20 0.8423
## Year1998 0.0604 0.0771 0.78 0.4332
## Year1999 -0.0738 0.0583 -1.27 0.2058
## Year2000 -0.0244 0.0731 -0.33 0.7390
## Year2001 -0.0441 0.0687 -0.64 0.5215
## Year2002 -0.0315 0.0601 -0.52 0.6004
## Year2003 -0.1495 0.0605 -2.47 0.0136 *
## Year2004 -0.0968 0.0560 -1.73 0.0842 .
## Year2005 -0.1315 0.0561 -2.35 0.0191 *
## Year2006 -0.0729 0.0567 -1.29 0.1981
```

```

## Year2007          -0.1430      0.0562   -2.55   0.0110 *
## Year2008          -0.0676      0.0589   -1.15   0.2508
## Year2009          -0.0368      0.0571   -0.64   0.5192
## Year2010          -0.1546      0.0591   -2.62   0.0089 **
## Year2011          -0.1089      0.0562   -1.94   0.0527 .
## Year2012          -0.1145      0.0577   -1.99   0.0472 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.017, Adjusted R-squared:  0.01
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 192 weights are ~= 1. The remaining 2218 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.012  0.864  0.952  0.895  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.15e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2475 -0.2755  0.0189  0.2714  1.7897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2350     0.0473  26.14 < 2e-16 ***
## LastAuthorFemale1 -0.0778     0.0198  -3.93 8.9e-05 ***
## Year1997          0.0125     0.0606   0.21  0.8367
## Year1998          0.0502     0.0771   0.65  0.5146
## Year1999         -0.0751     0.0585  -1.28  0.1998
## Year2000         -0.0301     0.0731  -0.41  0.6801
## Year2001         -0.0381     0.0687  -0.56  0.5789
## Year2002         -0.0283     0.0602  -0.47  0.6378
## Year2003         -0.1485     0.0602  -2.47  0.0137 *
## Year2004         -0.0949     0.0558  -1.70  0.0888 .
## Year2005         -0.1354     0.0562  -2.41  0.0160 *
## Year2006         -0.0683     0.0561  -1.22  0.2239
```

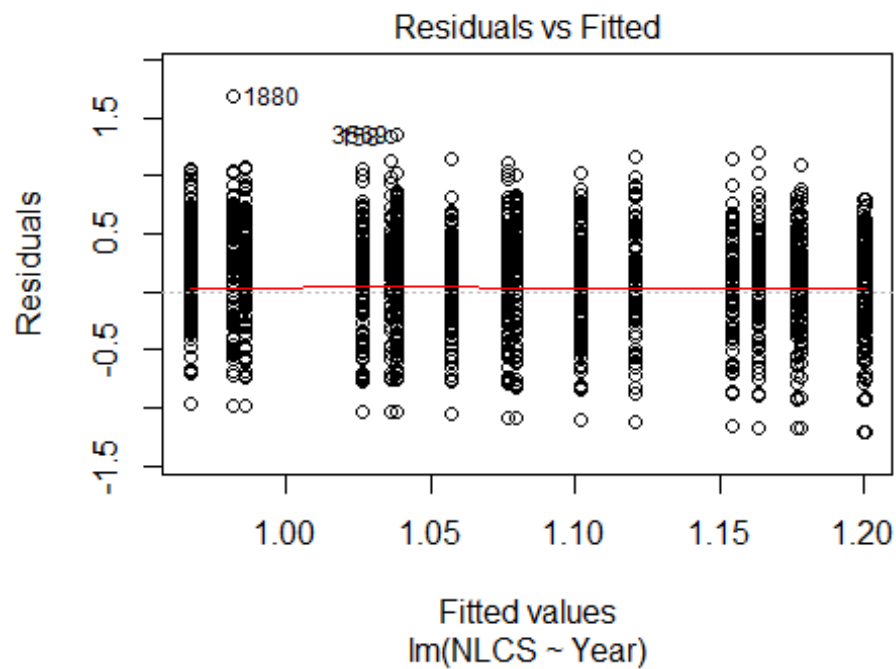
```

## Year2007          -0.1402      0.0560   -2.50   0.0124 *
## Year2008          -0.0677      0.0590   -1.15   0.2511
## Year2009          -0.0348      0.0568   -0.61   0.5398
## Year2010          -0.1530      0.0591   -2.59   0.0096 **
## Year2011          -0.1111      0.0561   -1.98   0.0476 *
## Year2012          -0.1135      0.0576   -1.97   0.0487 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.407
## Multiple R-squared:  0.0226, Adjusted R-squared:  0.0156
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 198 weights are ~= 1. The remaining 2212 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0138 0.8620 0.9520 0.8950 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.15e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2410"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2717"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 142 150 166 122 159 187 161 144 157 201 194 200 199 231 224
## 2011 2012
## 225 280
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 97 120 74 79 89 114 98 129 155 136 155 149 191 171
## 2011 2012

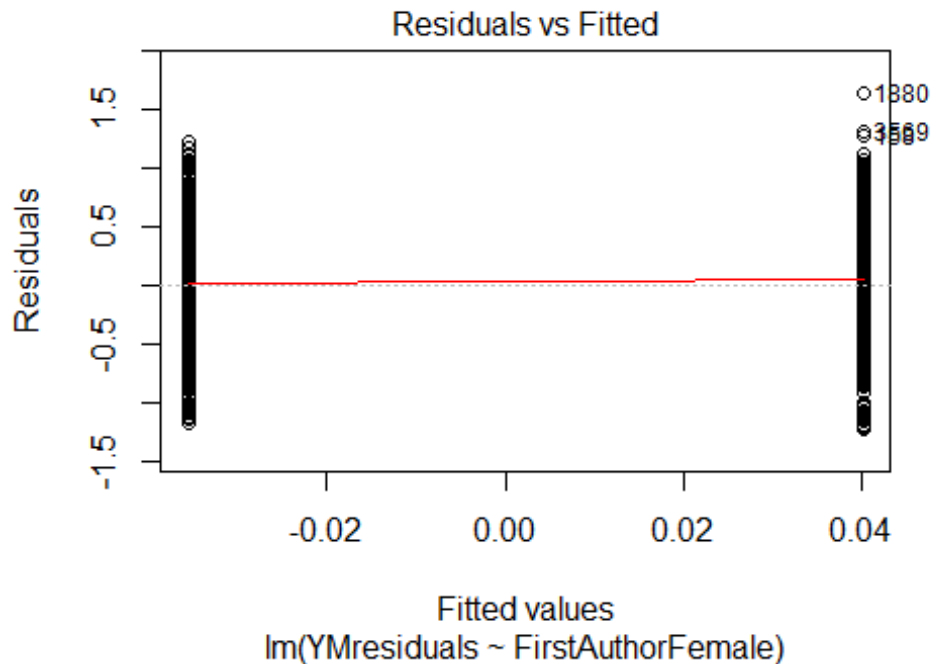
```



```
## 162 222
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 77 84 100 67 68 81 99 84 104 126 112 137 133 165 156
## 2011 2012
## 140 201
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```

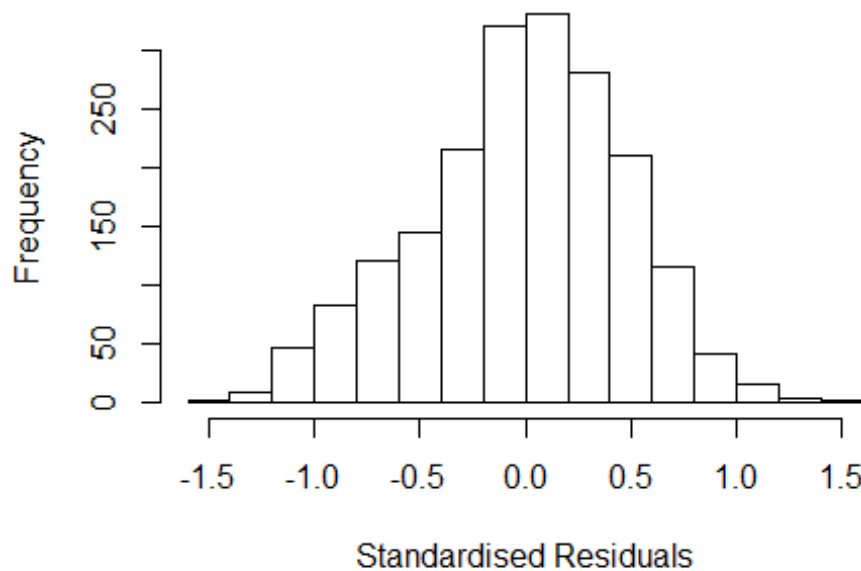


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.1169807254617"
## [1] "Male first author team size 2018 geometric mean: 4.39573995568986"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.05794440940954"
## [1] "Male last author team size 2018 geometric mean: 4.38808598481201"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3200, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.133 1      1.064
## LastAuthorFemale  1.096 1      1.047
## UniqueAuthors     1.313 4      1.035
## Year              1.380 16     1.010
```

## Residuals from first and last author and team size



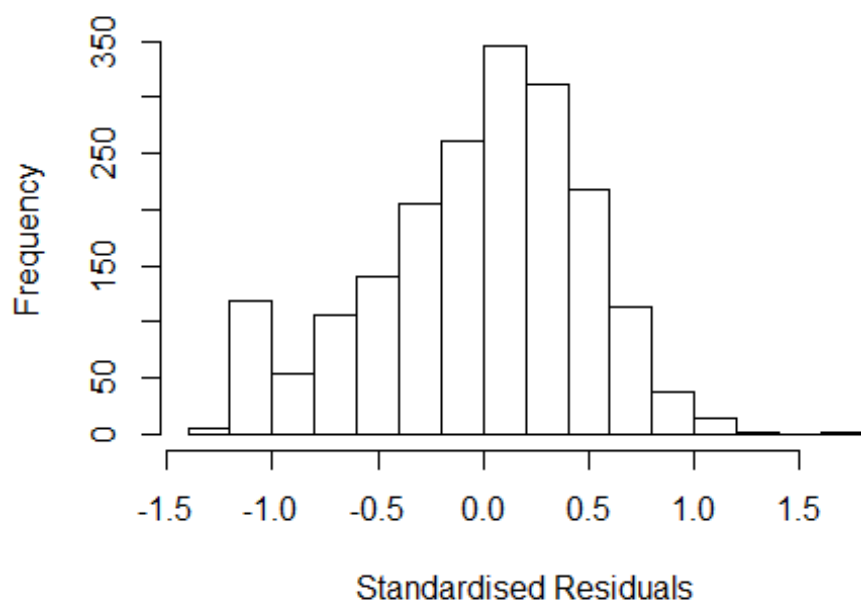
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.417 -0.329  0.020  0.318  1.477
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8010     0.0626   12.78 < 2e-16 ***
## FirstAuthorFemale1 0.0498     0.0240    2.08  0.038 *
## LastAuthorFemale1 0.0338     0.0252    1.34  0.180
## UniqueAuthors2    0.2010     0.0409    4.91 9.7e-07 ***
## UniqueAuthors3    0.3463     0.0427    8.10 9.5e-16 ***
## UniqueAuthors4    0.4044     0.0419    9.66 < 2e-16 ***
## UniqueAuthors5    0.4694     0.0368   12.75 < 2e-16 ***
## Year1997          0.1066     0.0743    1.43  0.151
## Year1998          0.0658     0.0740    0.89  0.374
## Year1999          0.1277     0.0790    1.62  0.106
```

```

## Year2000          0.0750      0.0761      0.99      0.325
## Year2001          0.0584      0.0795      0.73      0.463
## Year2002          0.0182      0.0748      0.24      0.808
## Year2003          0.0305      0.0781      0.39      0.697
## Year2004         -0.0286      0.0796     -0.36      0.720
## Year2005         -0.1231      0.0824     -1.49      0.135
## Year2006         -0.1348      0.0838     -1.61      0.108
## Year2007         -0.0261      0.0727     -0.36      0.720
## Year2008         -0.1361      0.0748     -1.82      0.069 .
## Year2009         -0.0796      0.0699     -1.14      0.255
## Year2010         -0.0837      0.0704     -1.19      0.235
## Year2011         -0.0831      0.0720     -1.16      0.248
## Year2012         -0.1051      0.0696     -1.51      0.131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.462
## Multiple R-squared:  0.129, Adjusted R-squared:  0.119
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 172 weights are ~= 1. The remaining 1762 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.286  0.856  0.946  0.900  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.102 1      1.050
## LastAuthorFemale  1.061 1      1.030
## Year              1.084 16      1.003

```

## Residuals from first and last author



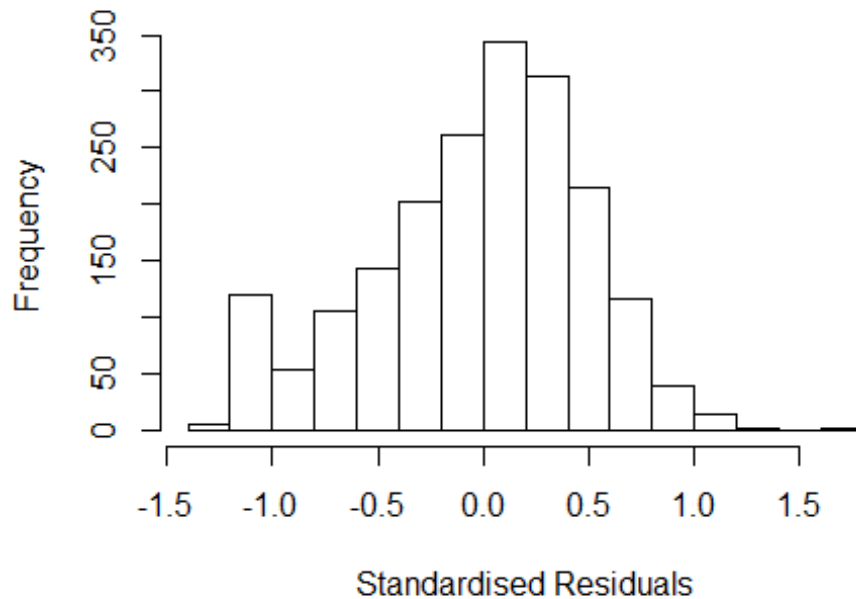
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2242 -0.3394 0.0464 0.3272 1.6212
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0061 0.0593 16.98 <2e-16 ***
## FirstAuthorFemale1 0.0723 0.0251 2.88 0.004 **
## LastAuthorFemale1 -0.0132 0.0264 -0.50 0.617
## Year1997 0.1531 0.0733 2.09 0.037 *
## Year1998 0.1452 0.0739 1.96 0.050 *
## Year1999 0.1123 0.0801 1.40 0.161
## Year2000 0.1574 0.0796 1.98 0.048 *
## Year2001 0.1590 0.0796 2.00 0.046 *
## Year2002 0.1070 0.0762 1.41 0.160
## Year2003 0.1388 0.0793 1.75 0.080 .
## Year2004 0.0456 0.0856 0.53 0.594
## Year2005 -0.0304 0.0844 -0.36 0.719
```

```

## Year2006          -0.0376      0.0906   -0.41    0.678
## Year2007           0.0861      0.0729    1.18    0.238
## Year2008          -0.0432      0.0777   -0.56    0.578
## Year2009           0.0117      0.0710    0.16    0.870
## Year2010           0.0427      0.0693    0.62    0.538
## Year2011           0.0662      0.0729    0.91    0.364
## Year2012           0.0304      0.0711    0.43    0.669
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.0204, Adjusted R-squared:  0.0112
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 154 weights are ~= 1. The remaining 1780 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.204  0.852  0.946  0.892  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1      1.031
## Year              1.062 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2304 -0.3384 0.0471 0.3269 1.6270
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00376 0.05920 16.96 <2e-16 ***
## FirstAuthorFemale1 0.06882 0.02473 2.78 0.0054 **
## Year1997 0.15329 0.07316 2.10 0.0363 *
## Year1998 0.14458 0.07386 1.96 0.0504 .
## Year1999 0.11309 0.08007 1.41 0.1580
## Year2000 0.15699 0.07939 1.98 0.0481 *
## Year2001 0.15786 0.07944 1.99 0.0471 *
## Year2002 0.10666 0.07603 1.40 0.1608
## Year2003 0.13847 0.07917 1.75 0.0805 .
## Year2004 0.04495 0.08545 0.53 0.5989
## Year2005 -0.03083 0.08429 -0.37 0.7146
## Year2006 -0.03757 0.09065 -0.41 0.6786
```

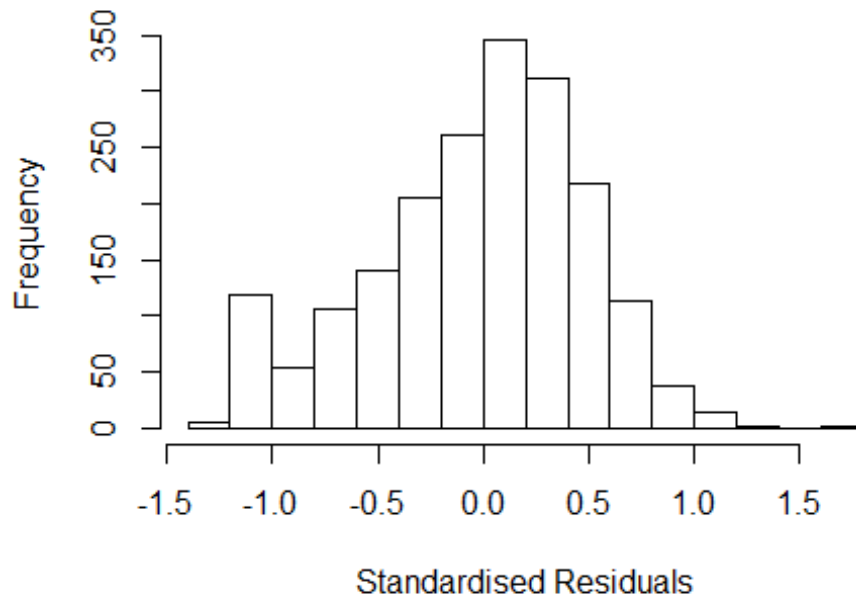
```

## Year2007          0.08541    0.07279    1.17    0.2408
## Year2008         -0.04338    0.07763   -0.56    0.5764
## Year2009          0.00989    0.07071    0.14    0.8888
## Year2010          0.04192    0.06921    0.61    0.5448
## Year2011          0.06506    0.07272    0.89    0.3711
## Year2012          0.02930    0.07095    0.41    0.6797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.0203, Adjusted R-squared:  0.0116
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 155 weights are ~= 1. The remaining 1779 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.201  0.852  0.945  0.892  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.012
## Year            1.023 16          1.001

```



## Residuals from last author



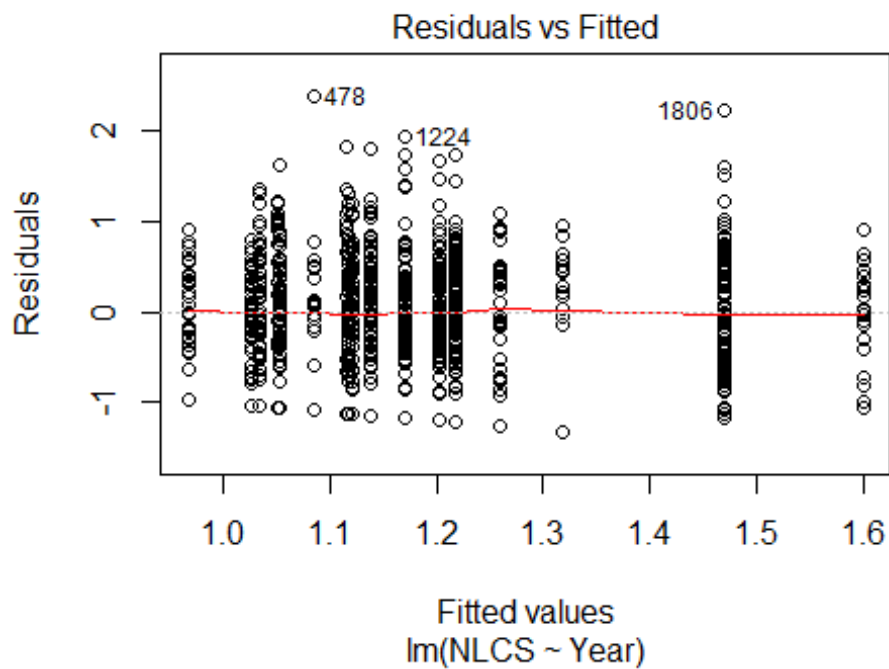
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1937 -0.3343  0.0466  0.3313  1.6636
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.02945    0.05762   17.86  <2e-16 ***
## LastAuthorFemale1 0.00724    0.02594    0.28   0.780
## Year1997        0.15142    0.07282    2.08   0.038 *
## Year1998        0.13870    0.07336    1.89   0.059 .
## Year1999        0.11373    0.07954    1.43   0.153
## Year2000        0.16421    0.07990    2.06   0.040 *
## Year2001        0.15044    0.07922    1.90   0.058 .
## Year2002        0.10605    0.07650    1.39   0.166
## Year2003        0.14277    0.07859    1.82   0.069 .
## Year2004        0.04461    0.08533    0.52   0.601
## Year2005       -0.03675    0.08384   -0.44   0.661
## Year2006       -0.03104    0.09028   -0.34   0.731
```

```

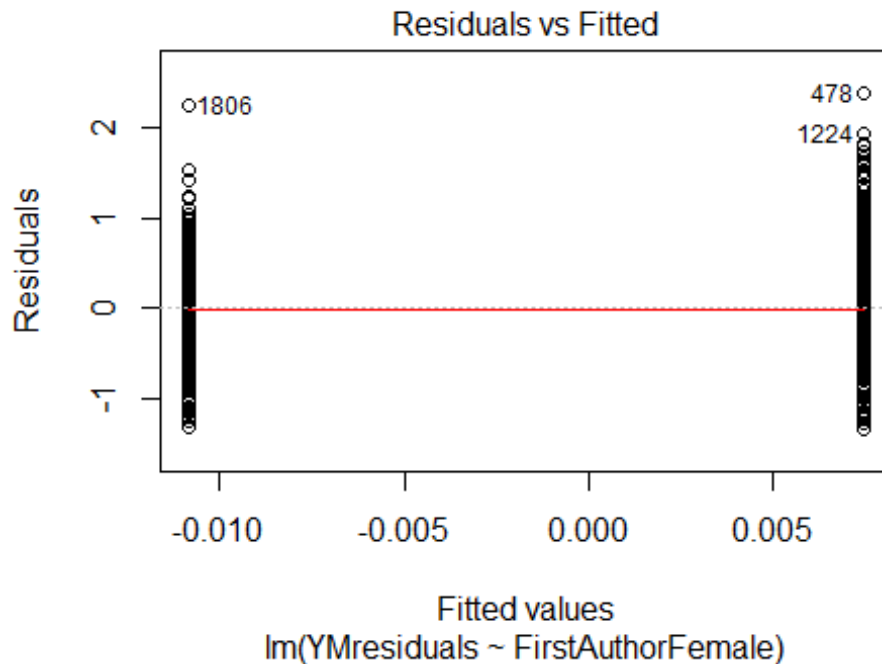
## Year2007      0.09036      0.07236      1.25      0.212
## Year2008     -0.04122      0.07717     -0.53      0.593
## Year2009      0.01523      0.07031      0.22      0.829
## Year2010      0.05065      0.06865      0.74      0.461
## Year2011      0.07037      0.07235      0.97      0.331
## Year2012      0.04431      0.07035      0.63      0.529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00731
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 159 weights are ~= 1. The remaining 1775 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.179  0.851  0.945  0.892  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1934"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2718"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   87   78   48   81   80   82   83   88   64   84  127  148   88   93  125
## 2011 2012
##  133  159
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   25   21   18   31   21   42   52   49   72   99  105   64   70   99
## 2011 2012

```

```
## 105 114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 23 18 16 27 20 41 48 45 65 91 92 59 59 93
## 2011 2012
## 90 101
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 52, df = 16, p-value = 1e-05
```

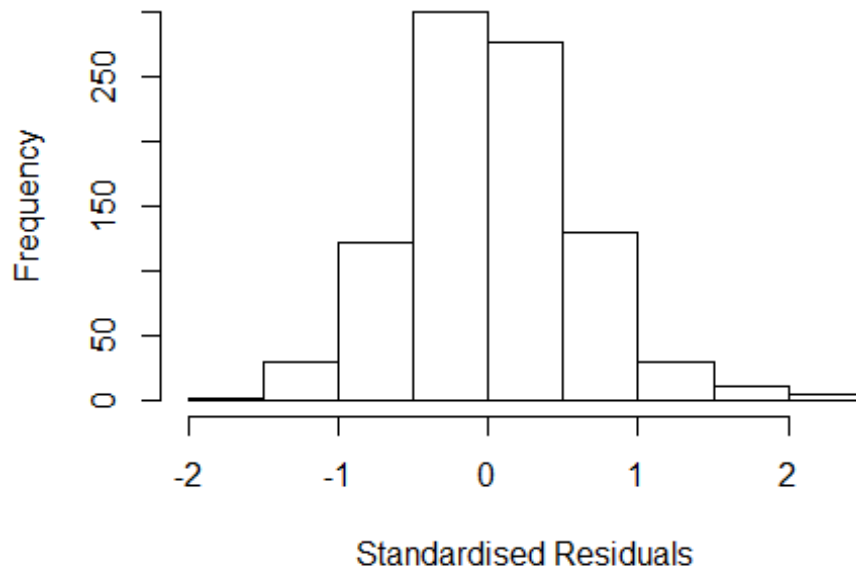


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.37306272468613"
## [1] "Male first author team size 2018 geometric mean: 4.2192980801933"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.26060943325362"
## [1] "Male last author team size 2018 geometric mean: 4.30297950449932"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.201 1          1.096
## LastAuthorFemale  1.246 1          1.116
## UniqueAuthors    1.556 4          1.057
## Year              1.752 16         1.018
```

## Residuals from first and last author and team size



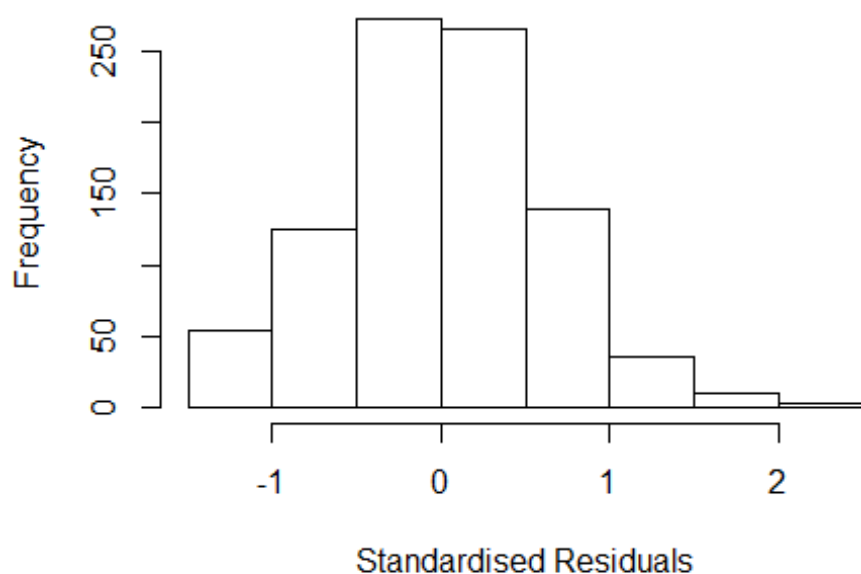
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.62596 -0.35520 -0.00204 0.39469 2.46697
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94008 0.42085 2.23 0.02575 *
## FirstAuthorFemale1 -0.01878 0.04177 -0.45 0.65307
## LastAuthorFemale1 -0.00239 0.04518 -0.05 0.95775
## UniqueAuthors2 0.27150 0.07370 3.68 0.00024 ***
## UniqueAuthors3 0.38081 0.07326 5.20 2.5e-07 ***
## UniqueAuthors4 0.39860 0.07116 5.60 2.8e-08 ***
## UniqueAuthors5 0.49390 0.06717 7.35 4.4e-13 ***
## Year1997 -0.24151 0.52194 -0.46 0.64368
## Year1998 0.46867 0.44002 1.07 0.28711
## Year1999 0.19197 0.47973 0.40 0.68913
```

```

## Year2000          0.00141    0.44820    0.00  0.99749
## Year2001         -0.20455    0.44253   -0.46  0.64403
## Year2002         -0.28271    0.43416   -0.65  0.51511
## Year2003         -0.20320    0.42826   -0.47  0.63528
## Year2004         -0.09099    0.43547   -0.21  0.83455
## Year2005         -0.22251    0.42961   -0.52  0.60463
## Year2006         -0.08438    0.43137   -0.20  0.84497
## Year2007         -0.14718    0.42848   -0.34  0.73131
## Year2008         -0.21727    0.43099   -0.50  0.61430
## Year2009         -0.14180    0.42914   -0.33  0.74115
## Year2010         -0.05550    0.42799   -0.13  0.89685
## Year2011          0.14673    0.42947    0.34  0.73270
## Year2012         -0.09286    0.42833   -0.22  0.82841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.135, Adjusted R-squared:  0.113
## Convergence in 28 IRWLS iterations
##
## Robustness weights:
## 63 weights are ~= 1. The remaining 842 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0101 0.8600 0.9520 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.133 1      1.064
## LastAuthorFemale  1.154 1      1.074
## Year              1.176 16      1.005

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43086 -0.37416 0.00238 0.41230 2.40986
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06735 0.46032 2.32 0.021 *
## FirstAuthorFemale1 -0.00210 0.04272 -0.05 0.961
## LastAuthorFemale1 -0.02484 0.04570 -0.54 0.587
## Year1997 -0.14028 0.55449 -0.25 0.800
## Year1998 0.59706 0.47898 1.25 0.213
## Year1999 0.36351 0.51089 0.71 0.477
## Year2000 0.19635 0.48087 0.41 0.683
## Year2001 -0.00320 0.47686 -0.01 0.995
## Year2002 -0.07887 0.46852 -0.17 0.866
## Year2003 -0.02667 0.46271 -0.06 0.954
## Year2004 0.09871 0.46911 0.21 0.833
## Year2005 -0.08849 0.46429 -0.19 0.849
```

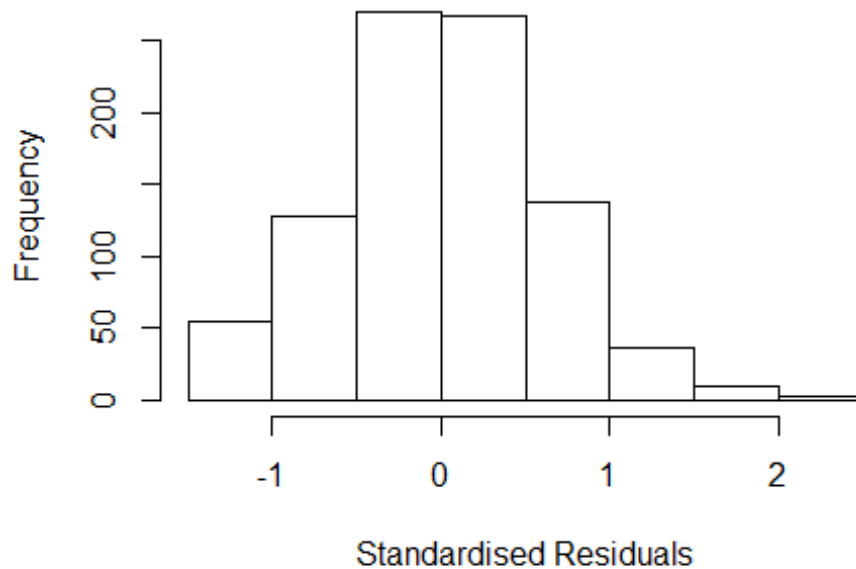
```

## Year2006          0.09536      0.46451      0.21      0.837
## Year2007          0.05533      0.46103      0.12      0.904
## Year2008          0.00456      0.46322      0.01      0.992
## Year2009          0.09693      0.46301      0.21      0.834
## Year2010          0.14981      0.46132      0.32      0.745
## Year2011          0.38621      0.46272      0.83      0.404
## Year2012          0.14933      0.46199      0.32      0.747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.579
## Multiple R-squared:  0.0595, Adjusted R-squared:  0.0403
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0442 0.8560 0.9460 0.8970 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.079 1      1.039
## Year              1.079 16      1.002

```



## Residuals from first author



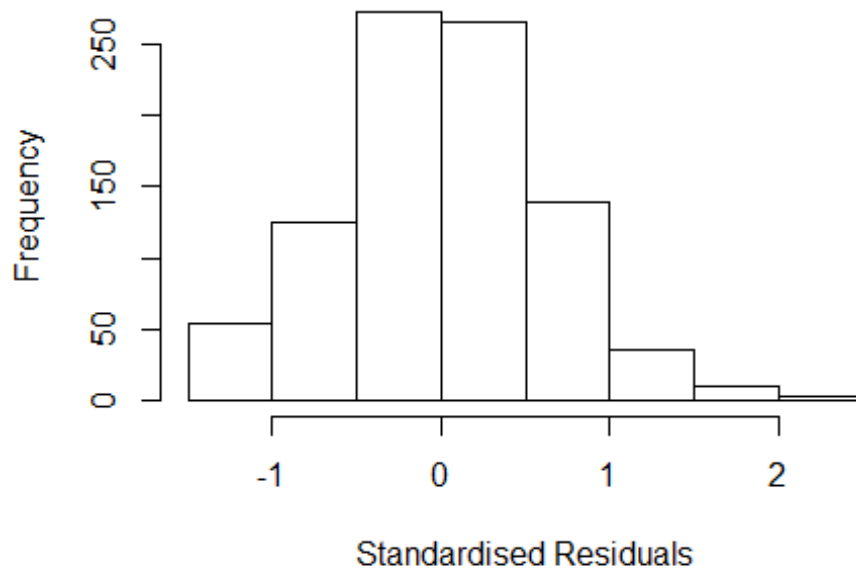
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.42452 -0.37016 0.00012 0.41080 2.41397
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.060968 0.455302 2.33 0.02 *
## FirstAuthorFemale1 -0.008986 0.041790 -0.22 0.83
## Year1997 -0.138554 0.553471 -0.25 0.80
## Year1998 0.601087 0.473830 1.27 0.20
## Year1999 0.363550 0.503240 0.72 0.47
## Year2000 0.195602 0.476760 0.41 0.68
## Year2001 -0.000934 0.473309 0.00 1.00
## Year2002 -0.078947 0.464534 -0.17 0.87
## Year2003 -0.027716 0.459041 -0.06 0.95
## Year2004 0.100954 0.464662 0.22 0.83
## Year2005 -0.089282 0.461387 -0.19 0.85
## Year2006 0.095234 0.460730 0.21 0.84
```

```

## Year2007          0.057167    0.457580    0.12    0.90
## Year2008          0.006691    0.459386    0.01    0.99
## Year2009          0.094912    0.459697    0.21    0.84
## Year2010          0.152081    0.457518    0.33    0.74
## Year2011          0.385190    0.460085    0.84    0.40
## Year2012          0.151394    0.457982    0.33    0.74
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.584
## Multiple R-squared:  0.0589, Adjusted R-squared:  0.0408
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 823 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0492 0.8600 0.9480 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.1 1          1.049
## Year              1.1 16          1.003

```

## Residuals from last author



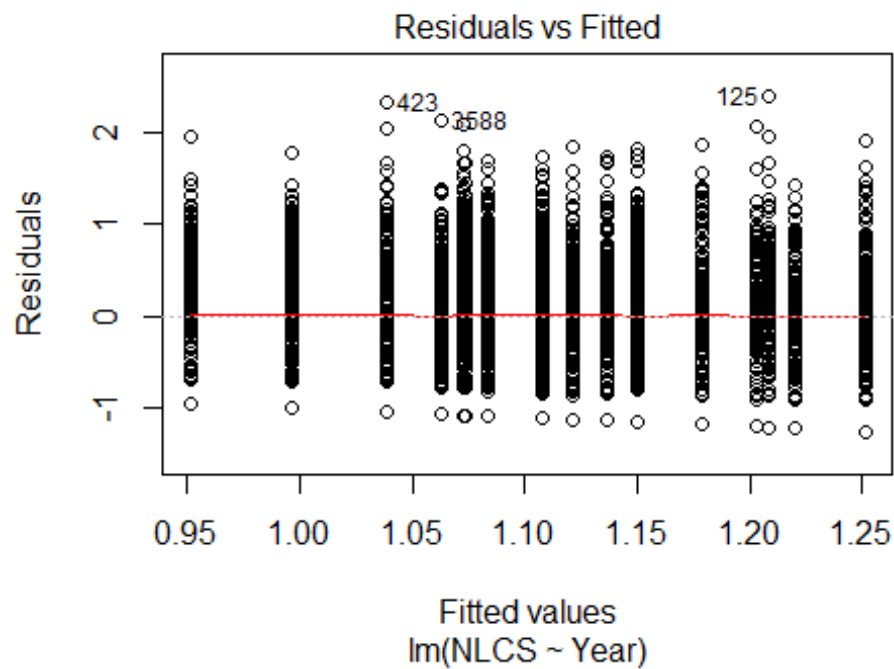
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.43132 -0.37333 0.00117 0.41311 2.41058
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06642 0.46080 2.31 0.021 *
## LastAuthorFemale1 -0.02546 0.04467 -0.57 0.569
## Year1997 -0.13942 0.55488 -0.25 0.802
## Year1998 0.59794 0.47961 1.25 0.213
## Year1999 0.36490 0.51080 0.71 0.475
## Year2000 0.19690 0.48170 0.41 0.683
## Year2001 -0.00300 0.47780 -0.01 0.995
## Year2002 -0.07864 0.46948 -0.17 0.867
## Year2003 -0.02646 0.46371 -0.06 0.955
## Year2004 0.09916 0.46999 0.21 0.833
## Year2005 -0.08843 0.46527 -0.19 0.849
## Year2006 0.09547 0.46545 0.21 0.838
```

```

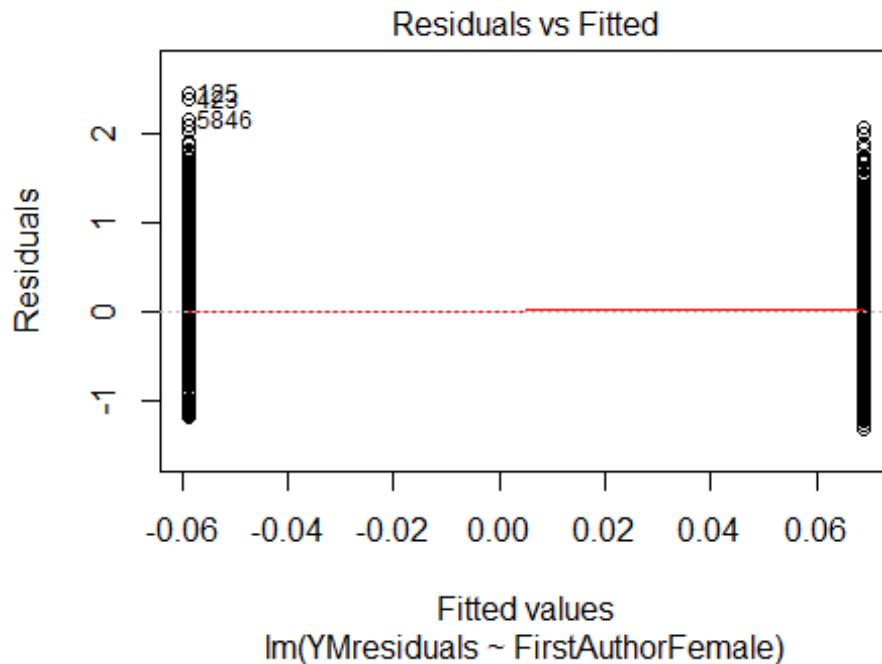
## Year2007      0.05541    0.46205    0.12    0.905
## Year2008      0.00469    0.46421    0.01    0.992
## Year2009      0.09722    0.46396    0.21    0.834
## Year2010      0.14991    0.46231    0.32    0.746
## Year2011      0.38631    0.46372    0.83    0.405
## Year2012      0.14957    0.46298    0.32    0.747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.0595, Adjusted R-squared:  0.0415
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 90 weights are ~= 1. The remaining 815 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0434 0.8550 0.9460 0.8970 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.10e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 905"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2719"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 279 297 302 294 334 324 289 323 315 389 408 548 629 594 642
## 2011 2012
## 714 687
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 185 181 190 183 172 201 238 233 272 318 437 470 444 444
## 2011 2012

```

```
## 562 541
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 190 179 173 184 176 166 188 219 207 253 280 411 436 405 412
## 2011 2012
## 513 506
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 20, df = 16, p-value = 0.2
```

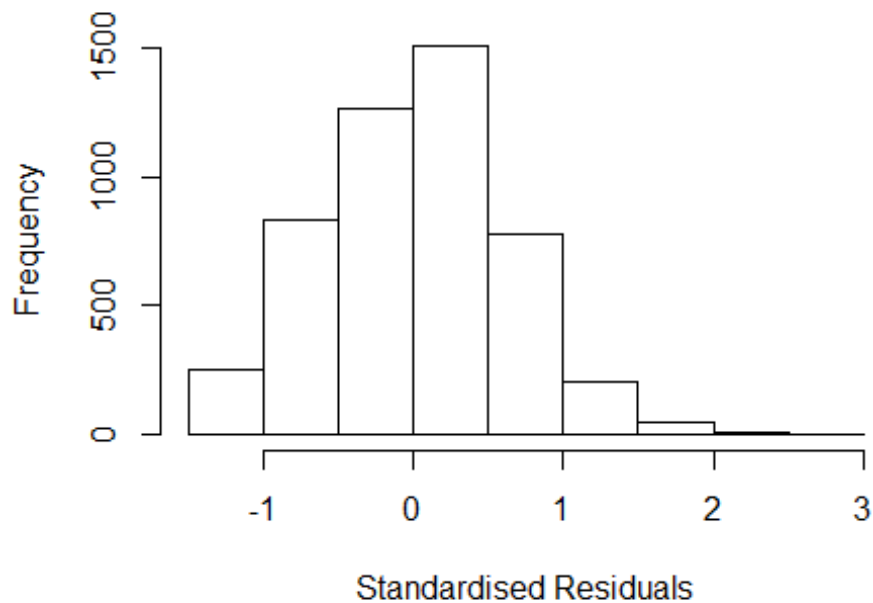


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.1, df = 1, p-value = 0.003
```



```
## [1] "Female first author team size 2018 geometric mean: 3.09227090141279"
## [1] "Male first author team size 2018 geometric mean: 2.81476166628378"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 58000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.93643417307522"
## [1] "Male last author team size 2018 geometric mean: 2.96856834074572"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 54000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.191 1      1.091
## LastAuthorFemale  1.158 1      1.076
## UniqueAuthors     1.157 4      1.018
## Year              1.140 16      1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 423 0030965809 3.36 1997      2719      1      2.531
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.471 -0.437  0.031  0.423  2.531
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.01683    0.05161   19.70 < 2e-16 ***
## FirstAuthorFemale1 0.07217    0.02012    3.59 0.00034 ***
## LastAuthorFemale1 0.03538    0.02007    1.76 0.07808 .
## UniqueAuthors2    0.21145    0.02684    7.88 4.1e-15 ***
## UniqueAuthors3    0.30512    0.02767   11.03 < 2e-16 ***
## UniqueAuthors4    0.34849    0.03178   10.97 < 2e-16 ***
## UniqueAuthors5    0.39124    0.02652   14.75 < 2e-16 ***
## Year1997         -0.18752    0.07228   -2.59 0.00951 **
## Year1998         -0.04295    0.07038   -0.61 0.54173
## Year1999          0.03644    0.07023    0.52 0.60383
```

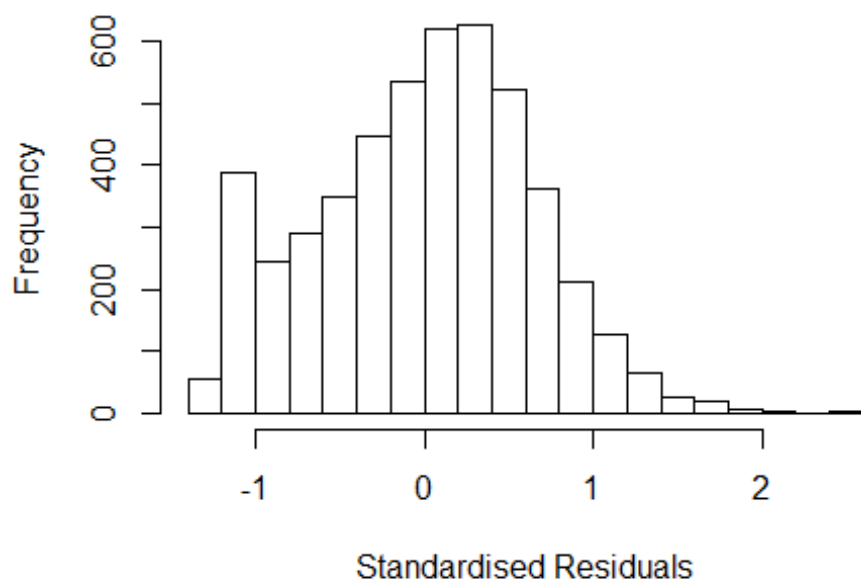
```

## Year2000      -0.00909      0.06849      -0.13      0.89443
## Year2001      -0.00645      0.07032      -0.09      0.92688
## Year2002      -0.10461      0.06746      -1.55      0.12107
## Year2003      -0.11664      0.06528      -1.79      0.07403 .
## Year2004      -0.17663      0.06972      -2.53      0.01133 *
## Year2005      -0.32383      0.06870      -4.71      2.5e-06 ***
## Year2006      -0.19173      0.06276      -3.05      0.00226 **
## Year2007      -0.18794      0.05900      -3.19      0.00146 **
## Year2008      -0.24306      0.05983      -4.06      4.9e-05 ***
## Year2009      -0.22126      0.06012      -3.68      0.00024 ***
## Year2010      -0.19269      0.05823      -3.31      0.00094 ***
## Year2011      -0.16386      0.05735      -2.86      0.00430 **
## Year2012      -0.13813      0.05720      -2.41      0.01578 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.633
## Multiple R-squared:  0.0783, Adjusted R-squared:  0.0741
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 391 weights are ~= 1. The remaining 4507 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0741 0.8610 0.9500 0.9110 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.04e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.145 1      1.070
## LastAuthorFemale 1.147 1      1.071
## Year      1.028 16      1.001

```



## Residuals from first and last author



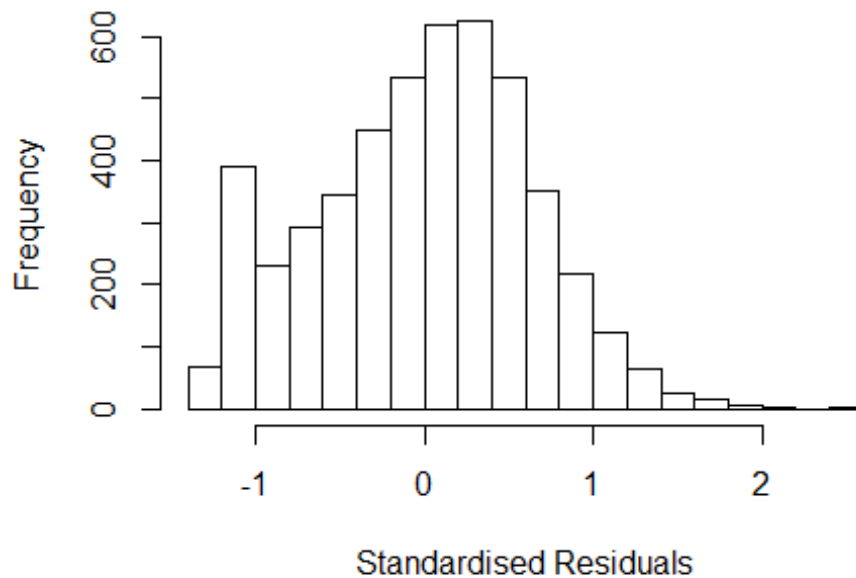
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3509 -0.4507  0.0421  0.4448  2.4663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1257    0.0507   22.19 < 2e-16 ***
## FirstAuthorFemale1  0.1285    0.0204    6.30 3.2e-10 ***
## LastAuthorFemale1  0.0269    0.0206    1.31 0.19178
## Year1997          -0.1759    0.0738   -2.38 0.01727 *
## Year1998          -0.0301    0.0718   -0.42 0.67525
## Year1999           0.0697    0.0704    0.99 0.32210
## Year2000           0.0502    0.0672    0.75 0.45513
## Year2001           0.0273    0.0690    0.40 0.69283
## Year2002          -0.0446    0.0683   -0.65 0.51415
## Year2003          -0.0508    0.0648   -0.78 0.43351
## Year2004          -0.1111    0.0684   -1.63 0.10423
## Year2005          -0.2601    0.0694   -3.75 0.00018 ***
```

```

## Year2006          -0.1185      0.0630   -1.88   0.05993 .
## Year2007          -0.1128      0.0596   -1.89   0.05844 .
## Year2008          -0.1821      0.0606   -3.01   0.00265 **
## Year2009          -0.1634      0.0602   -2.71   0.00670 **
## Year2010          -0.1196      0.0584   -2.05   0.04072 *
## Year2011          -0.0937      0.0576   -1.62   0.10432
## Year2012          -0.0571      0.0571   -1.00   0.31770
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.644
## Multiple R-squared:  0.0246, Adjusted R-squared:  0.021
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 415 weights are ~= 1. The remaining 4483 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.110  0.860  0.949  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling            cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.015 1      1.007
## Year              1.015 16      1.000

```

## Residuals from first author



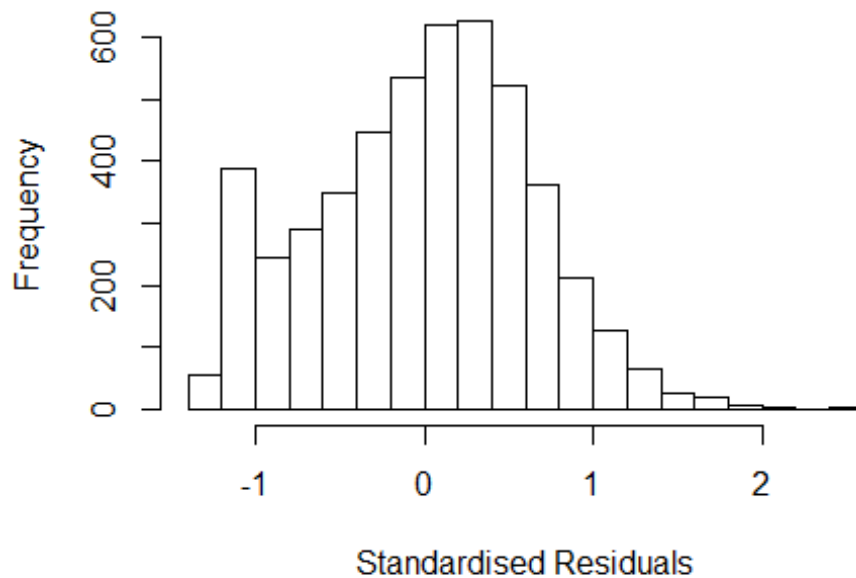
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3416 -0.4504  0.0378  0.4425  2.4618
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1302     0.0505   22.36 < 2e-16 ***
## FirstAuthorFemale1 0.1397     0.0193    7.25 4.8e-13 ***
## Year1997        -0.1757     0.0738   -2.38 0.01733 *
## Year1998        -0.0290     0.0717   -0.40 0.68606
## Year1999         0.0717     0.0704    1.02 0.30824
## Year2000         0.0516     0.0672    0.77 0.44277
## Year2001         0.0287     0.0690    0.42 0.67794
## Year2002        -0.0441     0.0684   -0.64 0.51956
## Year2003        -0.0504     0.0648   -0.78 0.43708
## Year2004        -0.1085     0.0683   -1.59 0.11207
## Year2005        -0.2595     0.0695   -3.73 0.00019 ***
## Year2006        -0.1184     0.0630   -1.88 0.06022 .
```

```

## Year2007          -0.1133      0.0596   -1.90  0.05737 .
## Year2008          -0.1805      0.0606   -2.98  0.00290 **
## Year2009          -0.1626      0.0603   -2.70  0.00700 **
## Year2010          -0.1180      0.0584   -2.02  0.04319 *
## Year2011          -0.0936      0.0577   -1.62  0.10445
## Year2012          -0.0552      0.0571   -0.97  0.33296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.644
## Multiple R-squared:  0.0242, Adjusted R-squared:  0.0208
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 431 weights are ~= 1. The remaining 4467 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.112  0.861  0.948  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1          1.008
## Year            1.016 16          1.000

```

## Residuals from last author



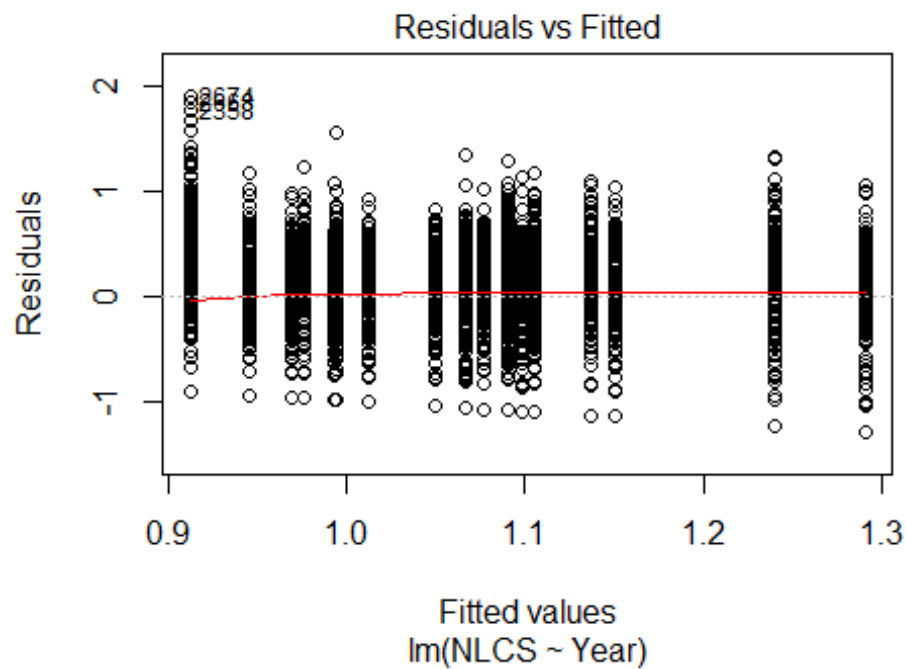
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3064 -0.4569 0.0428 0.4418 2.4246
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1674 0.0505 23.14 < 2e-16 ***
## LastAuthorFemale1 0.0821 0.0195 4.21 2.6e-05 ***
## Year1997 -0.1909 0.0746 -2.56 0.01050 *
## Year1998 -0.0399 0.0724 -0.55 0.58124
## Year1999 0.0569 0.0708 0.80 0.42179
## Year2000 0.0432 0.0676 0.64 0.52251
## Year2001 0.0177 0.0693 0.26 0.79787
## Year2002 -0.0474 0.0685 -0.69 0.48887
## Year2003 -0.0584 0.0649 -0.90 0.36825
## Year2004 -0.1185 0.0691 -1.71 0.08667 .
## Year2005 -0.2545 0.0696 -3.66 0.00026 ***
## Year2006 -0.1192 0.0633 -1.88 0.05981 .
```

```

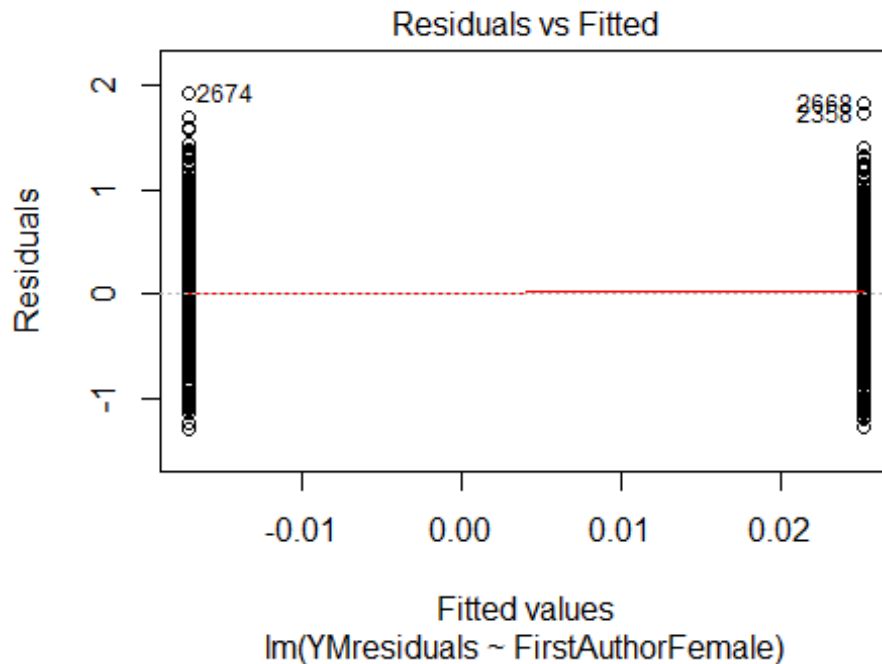
## Year2007          -0.1098      0.0601    -1.83   0.06775 .
## Year2008          -0.1905      0.0610    -3.12   0.00181 **
## Year2009          -0.1650      0.0605    -2.73   0.00641 **
## Year2010          -0.1171      0.0589    -1.99   0.04699 *
## Year2011          -0.0933      0.0582    -1.60   0.10881
## Year2012          -0.0623      0.0576    -1.08   0.27946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.648
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.0133
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 398 weights are ~= 1. The remaining 4500 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.131  0.865  0.949  0.910  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.04e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4898"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2720"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 538 524 433 402 636 452 453 330 322 301 344 375 374 321 304
## 2011 2012
## 357 360
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 127 142 112 138 240 174 252 189 185 181 203 212 219 202 219
## 2011 2012

```

```
## 244 209
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 114 126 103 121 202 149 212 154 156 134 174 186 191 167 188
## 2011 2012
## 217 178
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 220, df = 16, p-value <2e-16
```



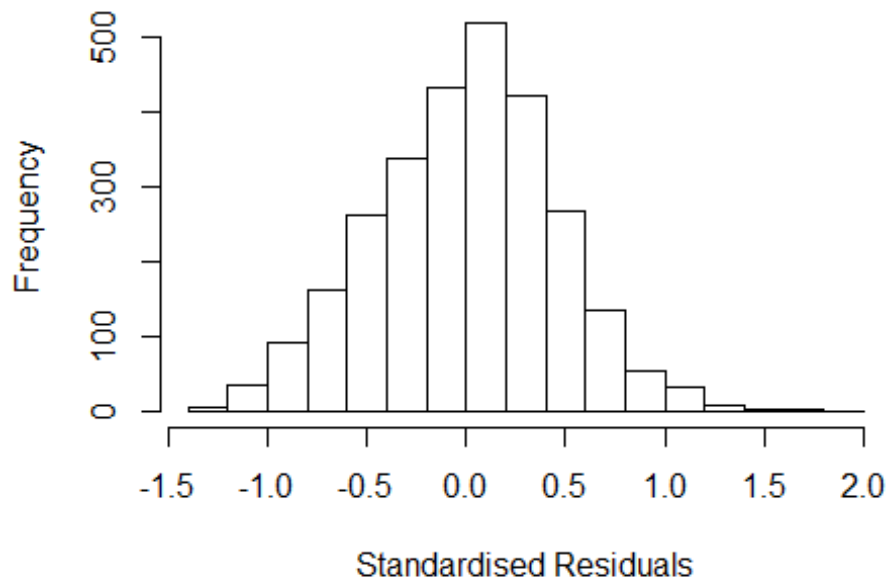
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.7, df = 1, p-value = 0.05
```



```
## [1] "Female first author team size 2018 geometric mean: 5.01073464476633"
## [1] "Male first author team size 2018 geometric mean: 4.89677360024674"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1800, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.86649987135012"
## [1] "Male last author team size 2018 geometric mean: 5.48880392810413"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.060 1          1.029
## LastAuthorFemale  1.027 1          1.013
## UniqueAuthors    1.235 4          1.027
## Year              1.255 16         1.007
```



## Residuals from first and last author and team size



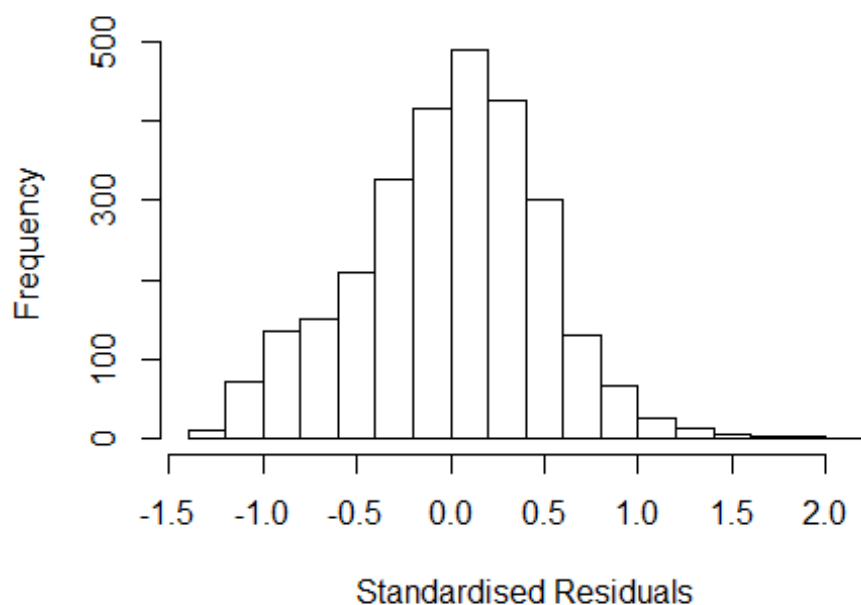
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3939 -0.3068 0.0211 0.3072 1.9495
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8924 0.0688 12.96 < 2e-16 ***
## FirstAuthorFemale1 0.0213 0.0184 1.16 0.24562
## LastAuthorFemale1 -0.0532 0.0215 -2.47 0.01357 *
## UniqueAuthors2 0.0549 0.0510 1.08 0.28183
## UniqueAuthors3 0.1766 0.0443 3.99 6.9e-05 ***
## UniqueAuthors4 0.2885 0.0452 6.38 2.1e-10 ***
## UniqueAuthors5 0.4801 0.0392 12.26 < 2e-16 ***
## Year1997 0.0984 0.0768 1.28 0.20031
## Year1998 0.1348 0.0734 1.84 0.06646 .
## Year1999 -0.0788 0.0701 -1.12 0.26120
```

```

## Year2000      -0.3234      0.0928      -3.49      0.00050 ***
## Year2001      -0.0690      0.0692      -1.00      0.31917
## Year2002      -0.1344      0.0621      -2.16      0.03062 *
## Year2003      -0.2651      0.0657      -4.03      5.6e-05 ***
## Year2004      -0.2378      0.0649      -3.67      0.00025 ***
## Year2005      -0.1863      0.0639      -2.91      0.00361 **
## Year2006      -0.2009      0.0663      -3.03      0.00246 **
## Year2007      -0.2348      0.0644      -3.64      0.00027 ***
## Year2008      -0.2086      0.0652      -3.20      0.00138 **
## Year2009      -0.1794      0.0639      -2.81      0.00503 **
## Year2010      -0.1263      0.0662      -1.91      0.05644 .
## Year2011      -0.1072      0.0650      -1.65      0.09949 .
## Year2012      -0.1130      0.0655      -1.72      0.08480 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.183, Adjusted R-squared:  0.176
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 232 weights are ~= 1. The remaining 2540 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0206 0.8680 0.9490 0.9040 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.61e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1 1.012
## LastAuthorFemale 1.016 1 1.008
## Year 1.040 16 1.001

```

## Residuals from first and last author



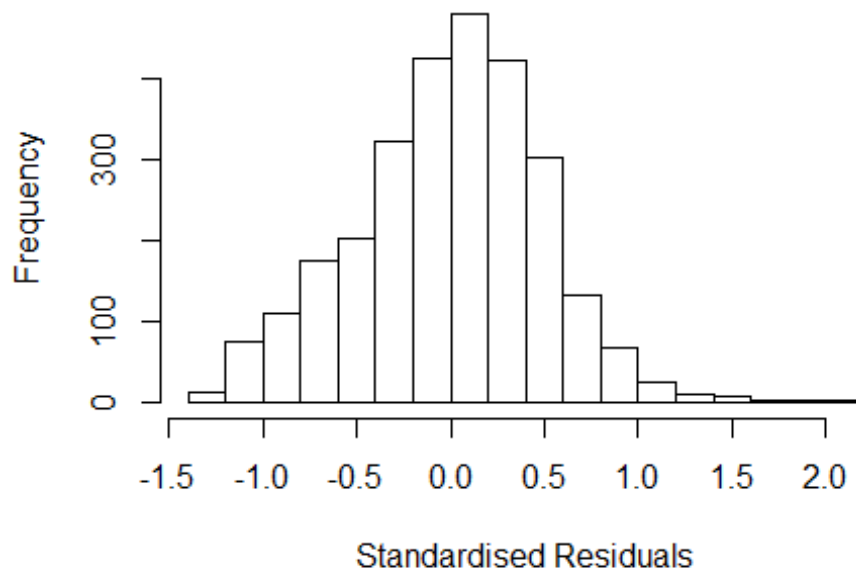
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3182 -0.3236 0.0338 0.3202 2.0057
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1743 0.0602 19.50 < 2e-16 ***
## FirstAuthorFemale1 0.0516 0.0193 2.67 0.00758 **
## LastAuthorFemale1 -0.0597 0.0232 -2.57 0.01014 *
## Year1997 0.1065 0.0795 1.34 0.18047
## Year1998 0.1439 0.0776 1.85 0.06375 .
## Year1999 -0.0725 0.0726 -1.00 0.31797
## Year2000 -0.3730 0.1058 -3.52 0.00043 ***
## Year2001 -0.0316 0.0725 -0.44 0.66230
## Year2002 -0.0695 0.0650 -1.07 0.28504
## Year2003 -0.2228 0.0691 -3.22 0.00128 **
## Year2004 -0.2077 0.0686 -3.03 0.00250 **
## Year2005 -0.1338 0.0680 -1.97 0.04918 *
```

```

## Year2006          -0.1819      0.0680    -2.68  0.00751 **
## Year2007          -0.2102      0.0690    -3.05  0.00233 **
## Year2008          -0.1992      0.0690    -2.89  0.00393 **
## Year2009          -0.1338      0.0682    -1.96  0.05001 .
## Year2010          -0.0589      0.0696    -0.85  0.39744
## Year2011          -0.0616      0.0684    -0.90  0.36739
## Year2012          -0.0664      0.0710    -0.94  0.34957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.464
## Multiple R-squared:  0.0579, Adjusted R-squared:  0.0518
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 220 weights are ~= 1. The remaining 2552 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0225 0.8590 0.9480 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.012
## Year      1.023 16      1.001

```

## Residuals from first author



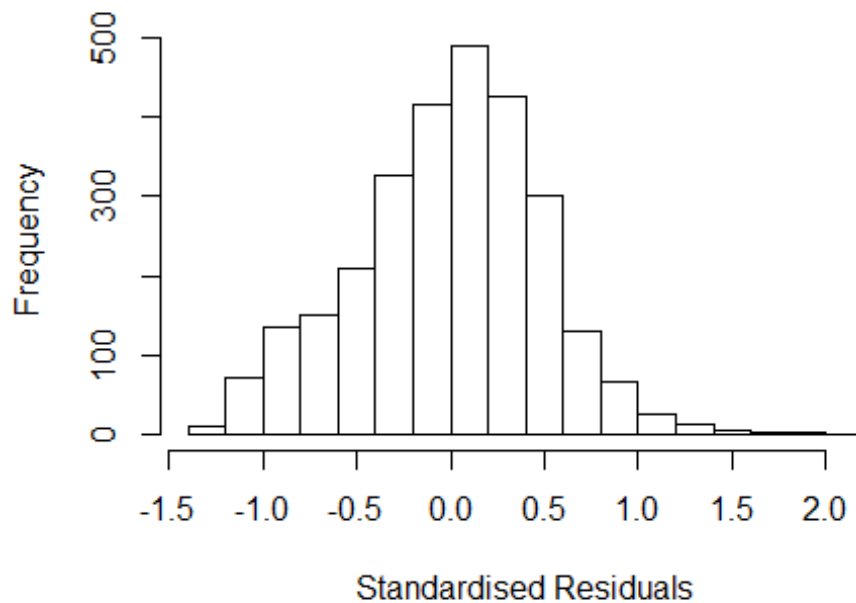
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3188 -0.3297 0.0322 0.3219 2.0196
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1621 0.0600 19.38 < 2e-16 ***
## FirstAuthorFemale1 0.0463 0.0194 2.38 0.01730 *
## Year1997 0.1104 0.0794 1.39 0.16454
## Year1998 0.1475 0.0774 1.91 0.05672 .
## Year1999 -0.0687 0.0724 -0.95 0.34253
## Year2000 -0.3747 0.1058 -3.54 0.00041 ***
## Year2001 -0.0282 0.0724 -0.39 0.69694
## Year2002 -0.0670 0.0651 -1.03 0.30391
## Year2003 -0.2212 0.0691 -3.20 0.00138 **
## Year2004 -0.2084 0.0687 -3.03 0.00245 **
## Year2005 -0.1322 0.0679 -1.95 0.05170 .
## Year2006 -0.1794 0.0680 -2.64 0.00841 **
```

```

## Year2007          -0.2109      0.0690   -3.05  0.00228 **
## Year2008          -0.2003      0.0692   -2.89  0.00383 **
## Year2009          -0.1347      0.0683   -1.97  0.04866 *
## Year2010          -0.0612      0.0697   -0.88  0.37992
## Year2011          -0.0618      0.0685   -0.90  0.36746
## Year2012          -0.0651      0.0711   -0.92  0.35936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0558, Adjusted R-squared:  0.05
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 210 weights are ~= 1. The remaining 2562 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0194 0.8600 0.9480 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1          1.008
## Year            1.017 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3409 -0.3246  0.0354  0.3211  1.9873
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1924    0.0594   20.09  < 2e-16 ***
## LastAuthorFemale1 -0.0533    0.0233   -2.28  0.02239 *
## Year1997          0.1091    0.0791    1.38  0.16798
## Year1998          0.1485    0.0770    1.93  0.05396 .
## Year1999         -0.0705    0.0722   -0.98  0.32910
## Year2000         -0.3727    0.1064   -3.50  0.00047 ***
## Year2001         -0.0292    0.0721   -0.40  0.68583
## Year2002         -0.0672    0.0648   -1.04  0.29965
## Year2003         -0.2206    0.0688   -3.20  0.00137 **
## Year2004         -0.2085    0.0684   -3.05  0.00231 **
## Year2005         -0.1320    0.0676   -1.95  0.05088 .
## Year2006         -0.1808    0.0676   -2.67  0.00756 **
```

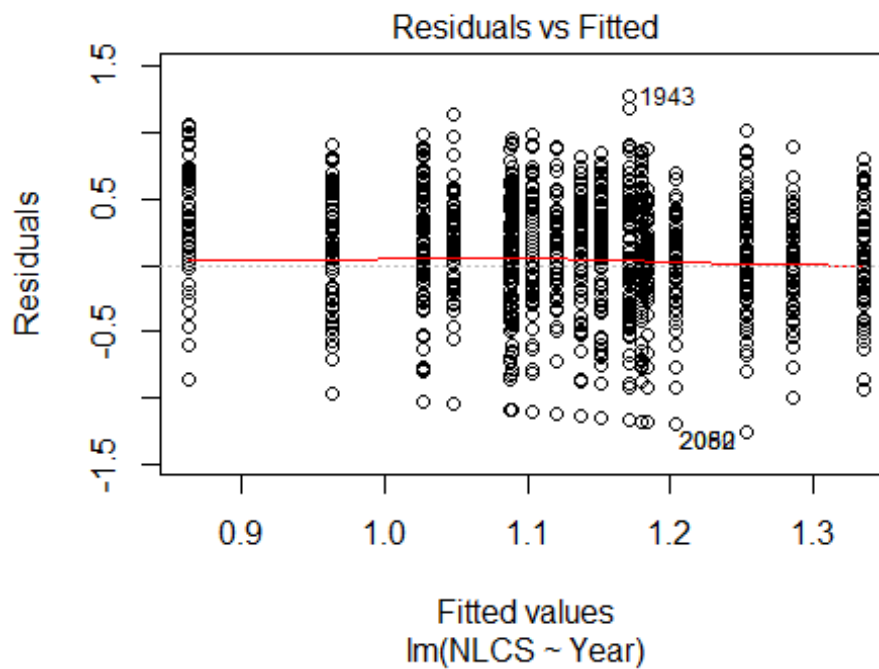
```

## Year2007          -0.2104      0.0689   -3.05  0.00228 **
## Year2008          -0.1971      0.0689   -2.86  0.00426 **
## Year2009          -0.1279      0.0679   -1.88  0.05976 .
## Year2010          -0.0574      0.0692   -0.83  0.40641
## Year2011          -0.0548      0.0681   -0.80  0.42130
## Year2012          -0.0664      0.0708   -0.94  0.34810
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0558, Adjusted R-squared:  0.05
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 204 weights are ~= 1. The remaining 2568 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0277 0.8620 0.9480 0.8970 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2772"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2721"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 126 126 127 136 132 155 115 105 123 95 119 141 136 131
## 2011 2012
## 123 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 65 66 66 68 54 87 98 77 65 82 57 76 97 95 97
## 2011 2012

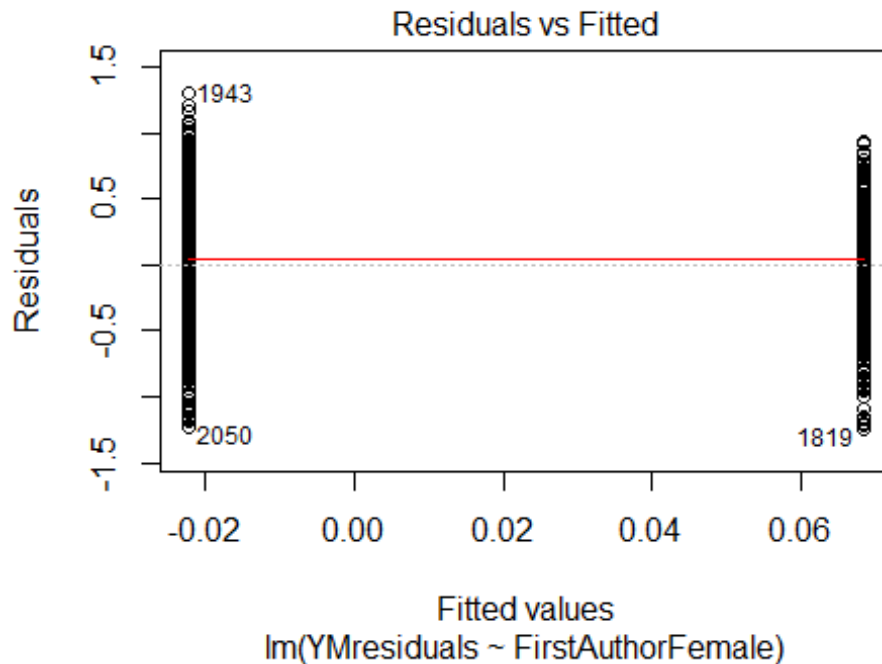
```



```
## 92 80
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 60 51 58 59 47 69 86 64 58 72 51 69 88 76 84
## 2011 2012
## 79 72
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 66, df = 16, p-value = 6e-08
```



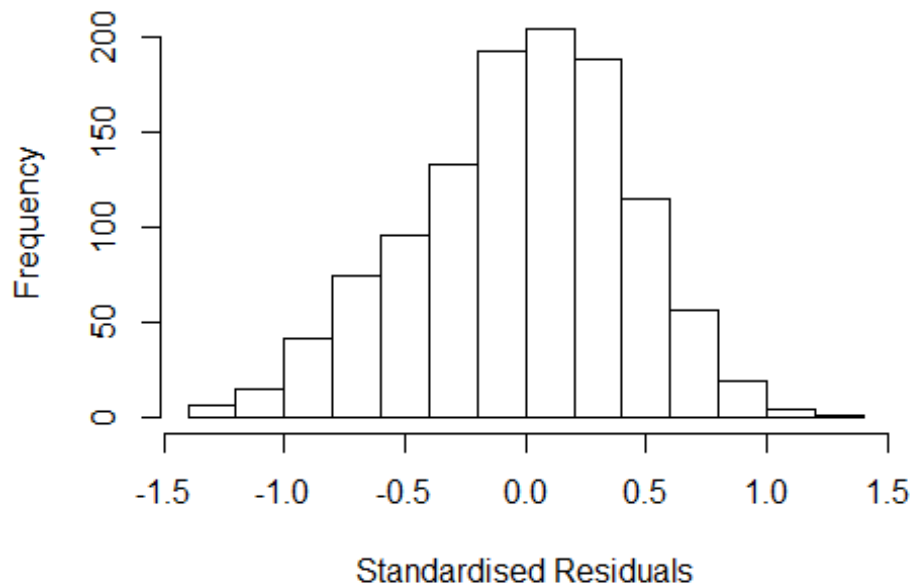
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.8, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 5.90804826364034"
## [1] "Male first author team size 2018 geometric mean: 5.41465610592805"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 620, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.40549657411827"
## [1] "Male last author team size 2018 geometric mean: 5.62872839698017"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 340, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.098	1	1.048
LastAuthorFemale	1.067	1	1.033
UniqueAuthors	1.442	4	1.047
Year	1.623	16	1.015

## Residuals from first and last author and team size



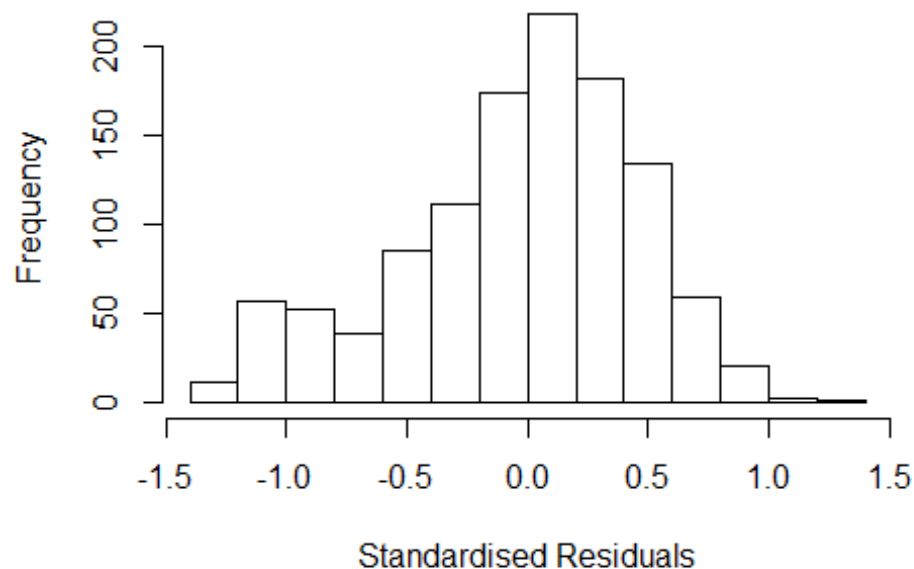
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3647 -0.2984  0.0188  0.3108  1.2624
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0269    0.0881   11.66 < 2e-16 ***
## FirstAuthorFemale1  0.0585    0.0322    1.82  0.06968 .
## LastAuthorFemale1  0.0373    0.0469    0.80  0.42657
## UniqueAuthors2   -0.1611    0.0964   -1.67  0.09505 .
## UniqueAuthors3    0.1226    0.0863    1.42  0.15583
## UniqueAuthors4    0.2168    0.0831    2.61  0.00921 **
## UniqueAuthors5    0.4465    0.0761    5.87  5.8e-09 ***
## Year1997         -0.0217    0.0738   -0.29  0.76881
## Year1998          0.0110    0.0734    0.15  0.88044
## Year1999         -0.1475    0.0708   -2.08  0.03751 *
```

```

## Year2000          -0.0968      0.0880    -1.10   0.27157
## Year2001          -0.2362      0.0819    -2.88   0.00402 **
## Year2002          -0.2214      0.0763    -2.90   0.00380 **
## Year2003          -0.2996      0.0797    -3.76   0.00018 ***
## Year2004          -0.3556      0.0952    -3.73   0.00020 ***
## Year2005          -0.1087      0.0782    -1.39   0.16464
## Year2006          -0.2256      0.0790    -2.86   0.00434 **
## Year2007          -0.1859      0.0758    -2.45   0.01438 *
## Year2008          -0.1566      0.0740    -2.12   0.03451 *
## Year2009          -0.0659      0.0771    -0.86   0.39273
## Year2010          -0.1807      0.0715    -2.53   0.01162 *
## Year2011          -0.1866      0.0822    -2.27   0.02330 *
## Year2012          -0.2007      0.0840    -2.39   0.01698 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.22,   Adjusted R-squared:  0.205
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1044 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.299  0.853  0.947   0.900  0.984   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           8.75e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.069 1           1.034
## LastAuthorFemale  1.056 1           1.028
## Year              1.126 16           1.004

```

## Residuals from first and last author



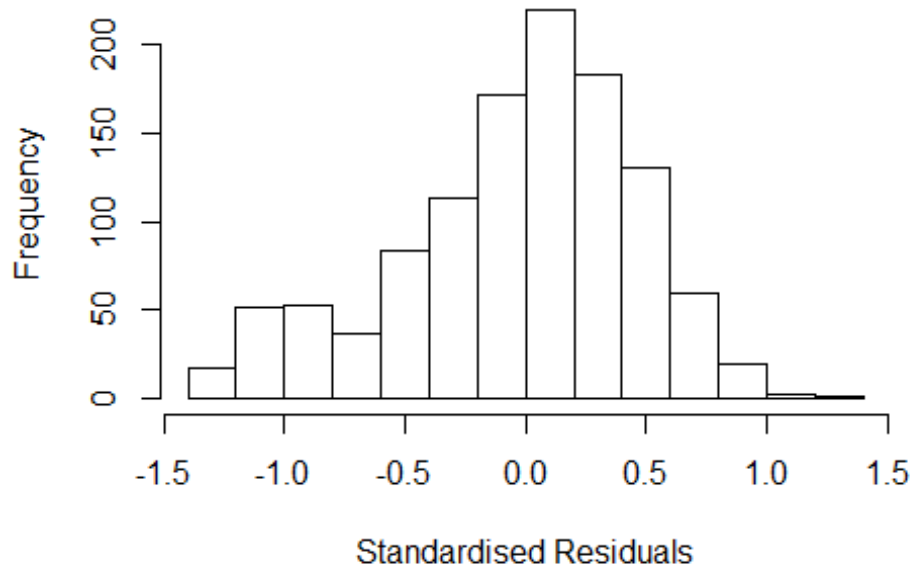
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2754 -0.3149 0.0387 0.3031 1.2624
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25904 0.04453 28.27 <2e-16 ***
## FirstAuthorFemale1 0.08650 0.03324 2.60 0.0094 **
## LastAuthorFemale1 0.02435 0.04971 0.49 0.6244
## Year1997 -0.02226 0.06293 -0.35 0.7236
## Year1998 0.04528 0.06990 0.65 0.5172
## Year1999 -0.06978 0.05967 -1.17 0.2425
## Year2000 -0.05933 0.09167 -0.65 0.5176
## Year2001 -0.26552 0.08122 -3.27 0.0011 **
## Year2002 -0.20168 0.08601 -2.34 0.0192 *
## Year2003 -0.22779 0.08353 -2.73 0.0065 **
## Year2004 -0.33602 0.12371 -2.72 0.0067 **
## Year2005 -0.07012 0.07370 -0.95 0.3416
```

```

## Year2006      -0.13859    0.07824   -1.77    0.0768 .
## Year2007      -0.10686    0.07739   -1.38    0.1676
## Year2008      -0.07444    0.07058   -1.05    0.2918
## Year2009       0.00714    0.06768    0.11    0.9160
## Year2010      -0.08222    0.06747   -1.22    0.2233
## Year2011      -0.10076    0.08187   -1.23    0.2187
## Year2012      -0.13854    0.07901   -1.75    0.0798 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.044, Adjusted R-squared:  0.0287
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 92 weights are ~= 1. The remaining 1051 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.394  0.855   0.947   0.888   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"    "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.067 1      1.033
## Year              1.067 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2777 -0.3163  0.0376  0.3058  1.2603
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2603     0.0446   28.27  <2e-16 ***
## FirstAuthorFemale1  0.0878     0.0332    2.64   0.0083 **
## Year1997         -0.0221     0.0630   -0.35   0.7256
## Year1998          0.0458     0.0699    0.66   0.5126
## Year1999         -0.0683     0.0596   -1.15   0.2518
## Year2000         -0.0592     0.0918   -0.65   0.5190
## Year2001         -0.2668     0.0811   -3.29   0.0010 **
## Year2002         -0.2008     0.0860   -2.33   0.0197 *
## Year2003         -0.2272     0.0834   -2.72   0.0066 **
## Year2004         -0.3348     0.1239   -2.70   0.0070 **
## Year2005         -0.0703     0.0736   -0.96   0.3395
## Year2006         -0.1387     0.0782   -1.77   0.0764 .
```

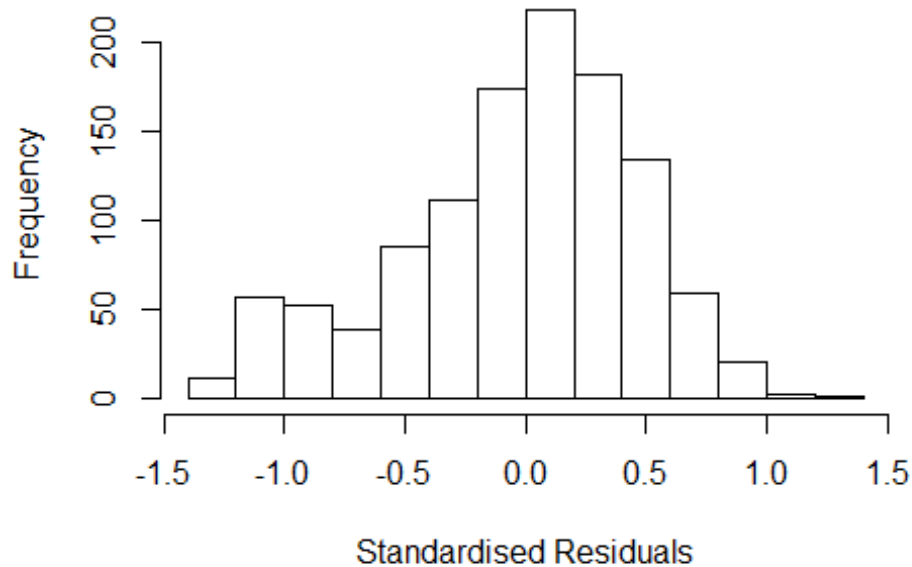
```

## Year2007          -0.1045      0.0770   -1.36   0.1751
## Year2008          -0.0736      0.0705   -1.04   0.2966
## Year2009           0.0075      0.0676    0.11   0.9117
## Year2010          -0.0799      0.0674   -1.18   0.2364
## Year2011          -0.1003      0.0818   -1.23   0.2204
## Year2012          -0.1357      0.0793   -1.71   0.0872 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.0438, Adjusted R-squared:  0.0293
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 1049 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.391  0.855   0.948   0.888   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.048 1          1.024
## Year            1.048 16          1.001

```



## Residuals from last author



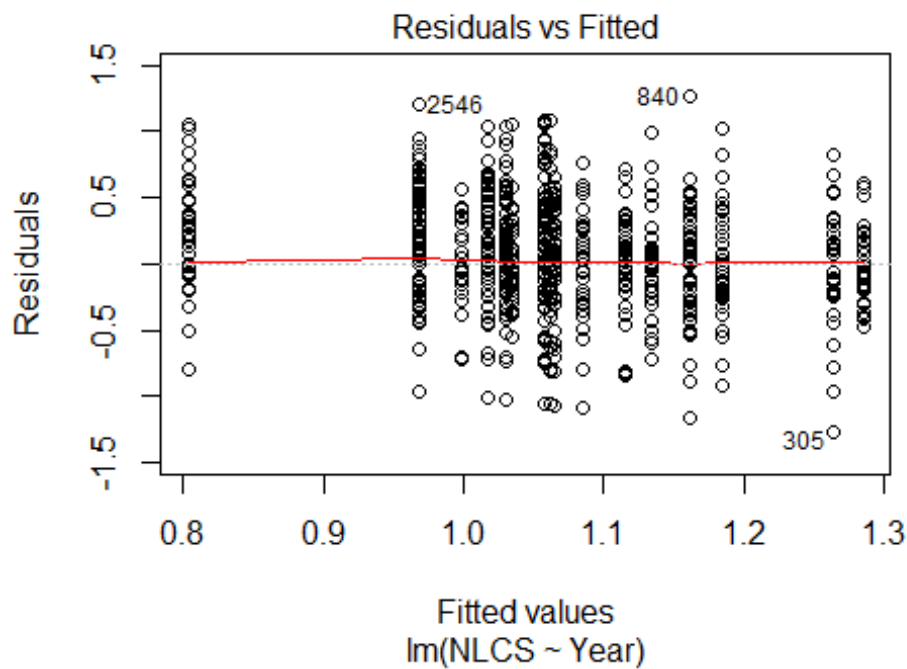
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2820 -0.3129  0.0435  0.3136  1.2369
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.27809    0.04341   29.44  <2e-16 ***
## LastAuthorFemale1  0.03480    0.04878    0.71  0.4758
## Year1997        -0.02125    0.06270   -0.34  0.7347
## Year1998         0.05474    0.06865    0.80  0.4254
## Year1999        -0.08063    0.05911   -1.36  0.1728
## Year2000        -0.05679    0.09185   -0.62  0.5365
## Year2001        -0.26153    0.08185   -3.20  0.0014 **
## Year2002        -0.20578    0.08648   -2.38  0.0175 *
## Year2003        -0.22828    0.08434   -2.71  0.0069 **
## Year2004        -0.34346    0.12352   -2.78  0.0055 **
## Year2005        -0.06069    0.07365   -0.82  0.4101
## Year2006        -0.12937    0.07841   -1.65  0.0992 .
```

```

## Year2007          -0.10377      0.07745    -1.34    0.1806
## Year2008          -0.06800      0.07074    -0.96    0.3366
## Year2009           0.00392      0.06670     0.06    0.9531
## Year2010          -0.08124      0.06770    -1.20    0.2304
## Year2011          -0.09639      0.08220    -1.17    0.2412
## Year2012          -0.12711      0.07841    -1.62    0.1053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.448
## Multiple R-squared:  0.0386, Adjusted R-squared:  0.0241
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1055 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.393  0.856  0.945  0.889  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1143"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2722"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 191 180 141 146 111 134 114 129 88 80 78 105 134 165 166
## 2011 2012
## 146 124
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33 38 40 40 18 39 32 47 26 36 29 40 57 103 89
## 2011 2012

```

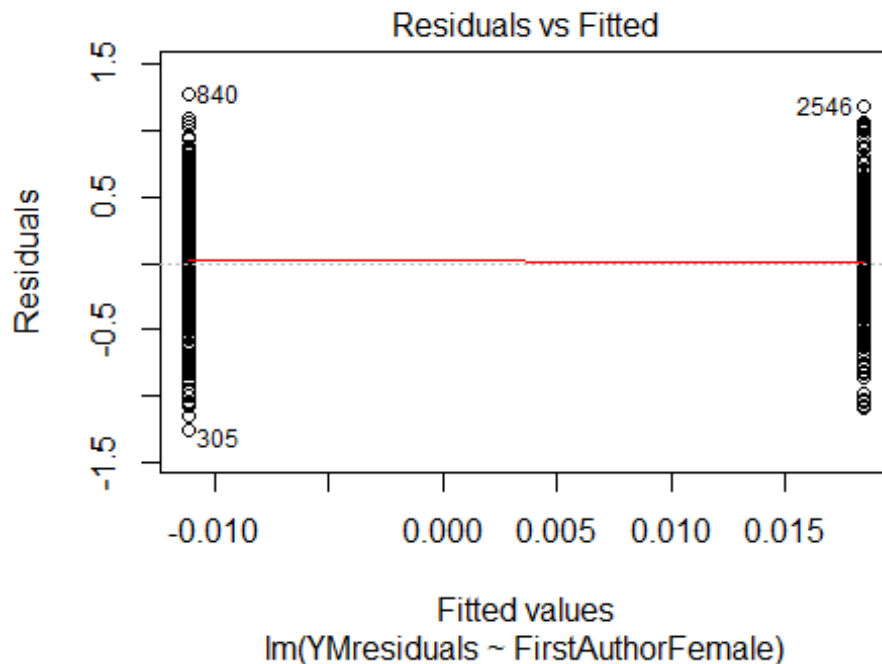
```
## 93 71
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 27 32 29 32 17 39 22 34 20 24 21 30 51 88 78
## 2011 2012
## 79 61
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 48, df = 16, p-value = 5e-05
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.7, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 4.51847561481269"
## [1] "Male first author team size 2018 geometric mean: 4.55036189106485"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

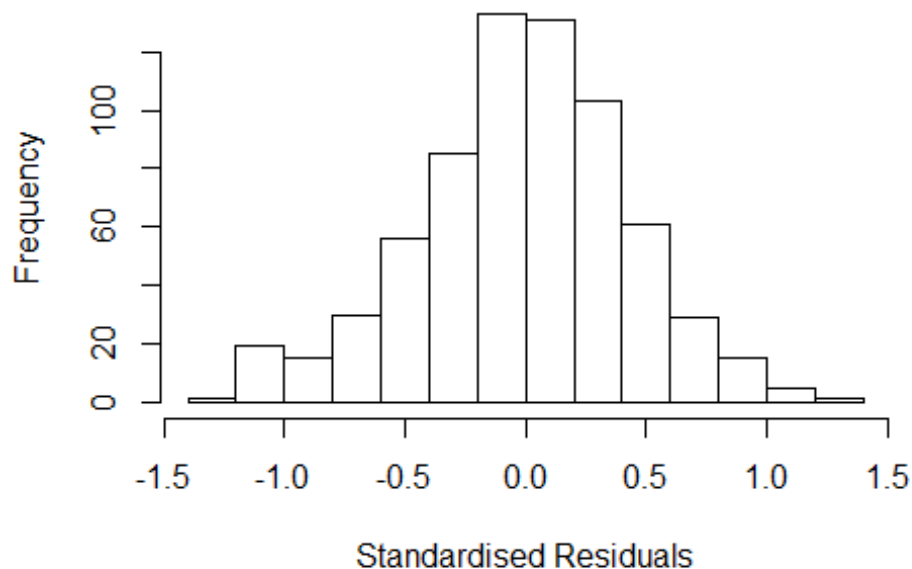
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 250, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.95720489293585"
## [1] "Male last author team size 2018 geometric mean: 4.62828566398529"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 96, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.201  1      1.096
## LastAuthorFemale  1.178  1      1.085
## UniqueAuthors    1.920  4      1.085
## Year              2.214 16      1.025
```

## Residuals from first and last author and team size



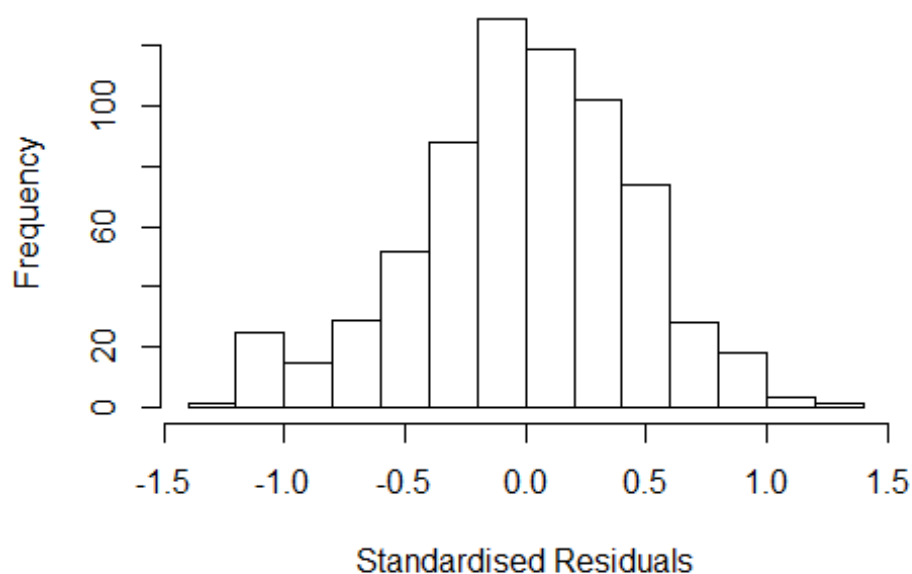
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3591 -0.2740  0.0105  0.2751  1.3113
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05768    0.07852   13.47 < 2e-16 ***
## FirstAuthorFemale1  0.01623    0.03740    0.43  0.66444
## LastAuthorFemale1 -0.00337    0.04241   -0.08  0.93664
## UniqueAuthors2     0.17996    0.06095    2.95  0.00326 **
## UniqueAuthors3     0.24312    0.06673    3.64  0.00029 ***
## UniqueAuthors4     0.26358    0.06808    3.87  0.00012 ***
## UniqueAuthors5     0.36201    0.05928    6.11  1.7e-09 ***
## Year1997          0.04122    0.09334    0.44  0.65896
## Year1998         -0.35021    0.12135   -2.89  0.00403 **
## Year1999         -0.09191    0.07300   -1.26  0.20846
```

```

## Year2000      -0.12498    0.11405   -1.10  0.27354
## Year2001      -0.06015    0.09511   -0.63  0.52734
## Year2002      -0.18793    0.09430   -1.99  0.04669 *
## Year2003      -0.22126    0.09541   -2.32  0.02069 *
## Year2004      -0.32515    0.08661   -3.75  0.00019 ***
## Year2005      -0.14371    0.09378   -1.53  0.12590
## Year2006      -0.09321    0.10689   -0.87  0.38352
## Year2007      -0.25695    0.09022   -2.85  0.00454 **
## Year2008      -0.40118    0.09649   -4.16  3.6e-05 ***
## Year2009      -0.21232    0.07962   -2.67  0.00785 **
## Year2010      -0.24312    0.07508   -3.24  0.00126 **
## Year2011      -0.26921    0.08804   -3.06  0.00232 **
## Year2012      -0.23294    0.09209   -2.53  0.01166 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.418
## Multiple R-squared:  0.12,   Adjusted R-squared:  0.0906
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 624 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.269  0.869  0.952  0.898  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.46e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.124 1          1.060
## LastAuthorFemale 1.174 1          1.084
## Year 1.239 16          1.007

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29634 -0.28058  0.00327  0.30360  1.25249
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.27753    0.05473   23.34 < 2e-16 ***
## FirstAuthorFemale1  0.05303    0.03805    1.39  0.16387
## LastAuthorFemale1 -0.00925    0.04427   -0.21  0.83451
## Year1997         0.02806    0.08891    0.32  0.75237
## Year1998        -0.30992    0.12330   -2.51  0.01219 *
## Year1999        -0.11218    0.07525   -1.49  0.13651
## Year2000        -0.10601    0.10833   -0.98  0.32812
## Year2001        -0.12352    0.08875   -1.39  0.16449
## Year2002        -0.23795    0.08630   -2.76  0.00599 **
## Year2003        -0.22446    0.09244   -2.43  0.01544 *
## Year2004        -0.29234    0.08464   -3.45  0.00059 ***
## Year2005        -0.13201    0.08442   -1.56  0.11838
```

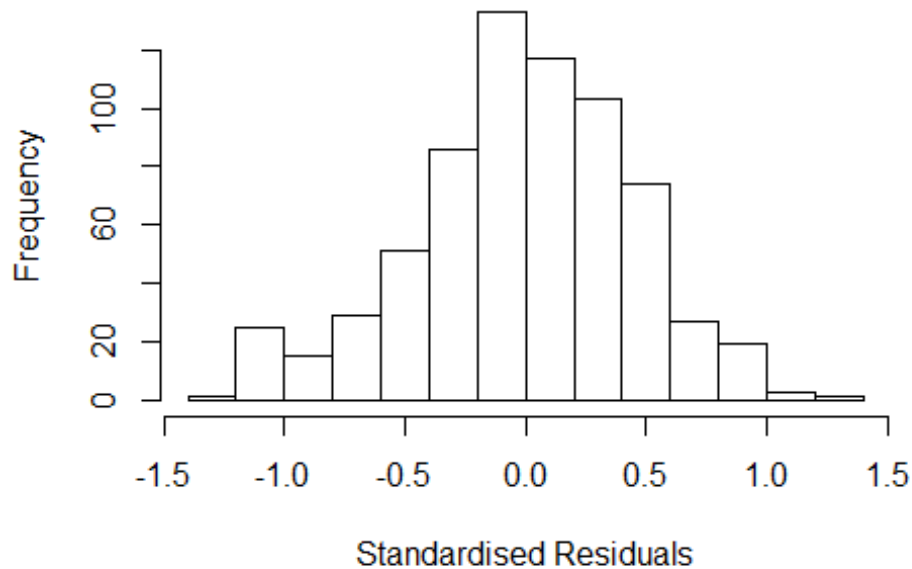
```

## Year2006      -0.02954    0.09837   -0.30  0.76404
## Year2007      -0.20711    0.08575   -2.42  0.01598 *
## Year2008      -0.42937    0.09980   -4.30  1.9e-05 ***
## Year2009      -0.19669    0.07967   -2.47  0.01380 *
## Year2010      -0.24860    0.07539   -3.30  0.00103 **
## Year2011      -0.27440    0.08682   -3.16  0.00165 **
## Year2012      -0.22575    0.08831   -2.56  0.01080 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0578, Adjusted R-squared:  0.0323
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 623 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.351  0.869  0.949  0.899  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.46e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1      1.041
## Year      1.084 16      1.003

```



## Residuals from first author



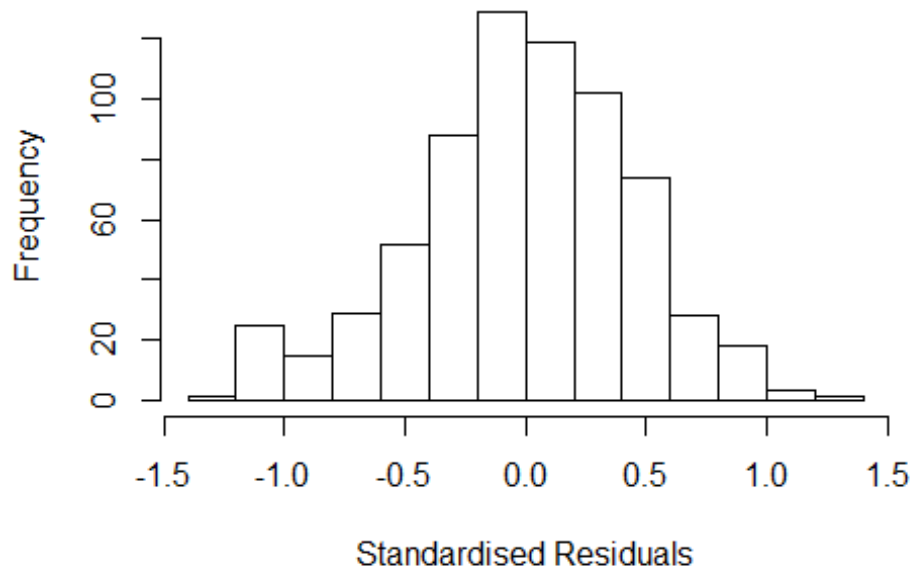
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30538 -0.28078 0.00248 0.30362 1.25275
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2758 0.0546 23.37 < 2e-16 ***
## FirstAuthorFemale1 0.0513 0.0374 1.37 0.17067
## Year1997 0.0296 0.0882 0.34 0.73742
## Year1998 -0.3099 0.1236 -2.51 0.01244 *
## Year1999 -0.1120 0.0755 -1.48 0.13854
## Year2000 -0.1045 0.1085 -0.96 0.33555
## Year2001 -0.1244 0.0886 -1.40 0.16084
## Year2002 -0.2366 0.0865 -2.73 0.00643 **
## Year2003 -0.2235 0.0925 -2.42 0.01599 *
## Year2004 -0.2907 0.0845 -3.44 0.00062 ***
## Year2005 -0.1309 0.0847 -1.55 0.12250
## Year2006 -0.0276 0.0983 -0.28 0.77912
```

```

## Year2007          -0.2070      0.0857   -2.41  0.01603 *
## Year2008          -0.4296      0.1000   -4.30   2e-05 ***
## Year2009          -0.1971      0.0797   -2.47  0.01362 *
## Year2010          -0.2493      0.0752   -3.32  0.00096 ***
## Year2011          -0.2742      0.0869   -3.16  0.00167 **
## Year2012          -0.2253      0.0885   -2.55  0.01112 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0579, Adjusted R-squared:  0.0338
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 623 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.342  0.867   0.949   0.898   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.46e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.132 1          1.064
## Year            1.132 16          1.004

```

## Residuals from last author



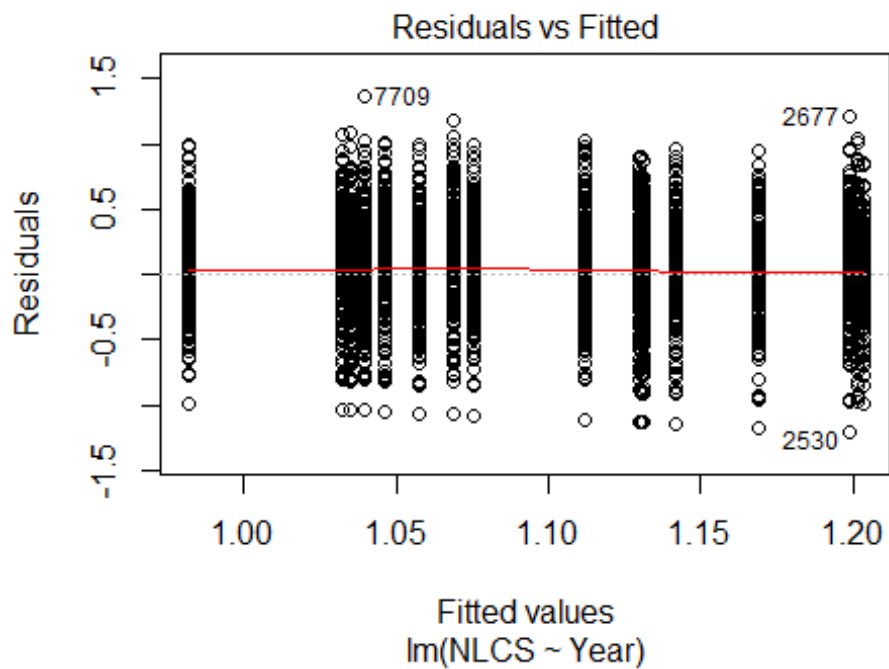
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33010 -0.28021 0.00159 0.30144 1.23593
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29197 0.05318 24.29 < 2e-16 ***
## LastAuthorFemale1 0.00388 0.04356 0.09 0.92910
## Year1997 0.03426 0.08736 0.39 0.69506
## Year1998 -0.30667 0.12322 -2.49 0.01306 *
## Year1999 -0.11454 0.07589 -1.51 0.13167
## Year2000 -0.10390 0.10839 -0.96 0.33813
## Year2001 -0.11864 0.08749 -1.36 0.17554
## Year2002 -0.23376 0.08587 -2.72 0.00665 **
## Year2003 -0.21178 0.09097 -2.33 0.02021 *
## Year2004 -0.29012 0.08415 -3.45 0.00060 ***
## Year2005 -0.12792 0.08244 -1.55 0.12123
## Year2006 -0.02499 0.09605 -0.26 0.79479
```

```

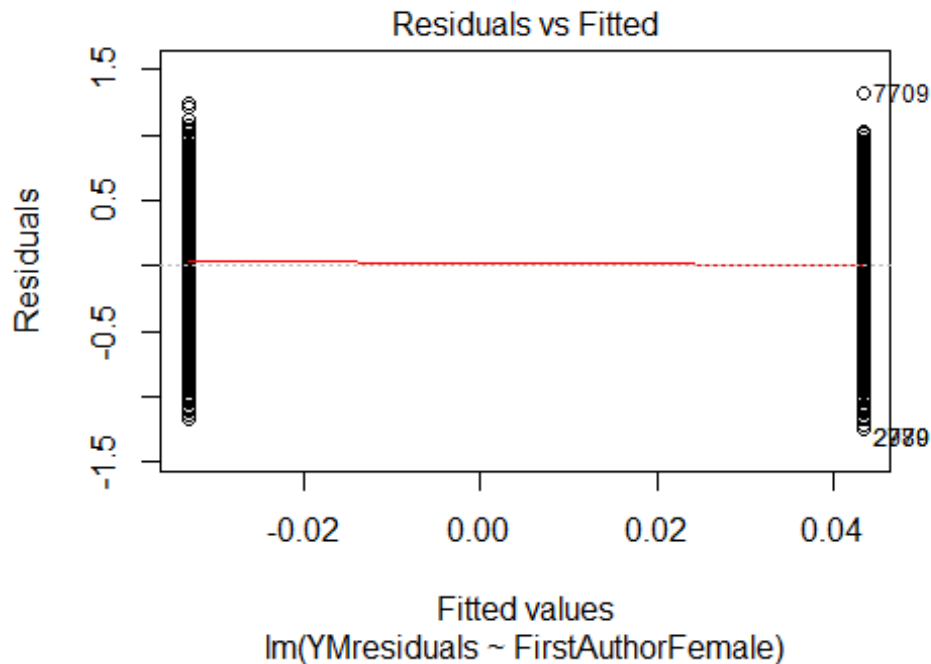
## Year2007          -0.20565      0.08557    -2.40  0.01652 *
## Year2008          -0.43059      0.10003    -4.30  1.9e-05 ***
## Year2009          -0.19838      0.07935    -2.50  0.01266 *
## Year2010          -0.24914      0.07525    -3.31  0.00098 ***
## Year2011          -0.26686      0.08726    -3.06  0.00232 **
## Year2012          -0.22723      0.08865    -2.56  0.01059 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.0556, Adjusted R-squared:  0.0315
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 622 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.324  0.867  0.953  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.46e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 684"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2723"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 485 463 485 477 452 476 504 394 396 369 350 382 406 373 369
## 2011 2012
## 376 393
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 198 177 219 216 187 149 280 228 213 216 186 240 237 218 236
## 2011 2012

```

```
## 251 267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 176 155 193 185 169 130 244 191 182 176 157 212 203 189 190
## 2011 2012
## 220 224
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 59, df = 16, p-value = 8e-07
```

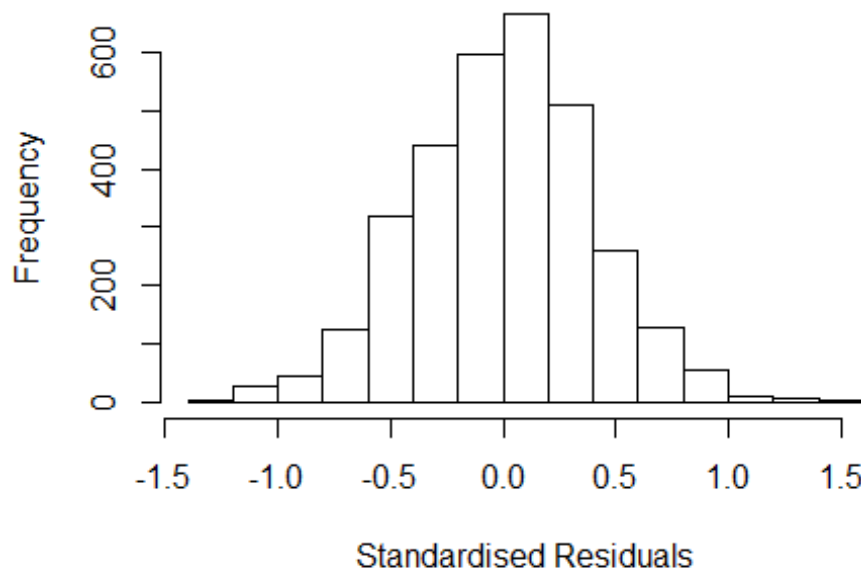


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 49, df = 1, p-value = 2e-12
```



```
## [1] "Female first author team size 2018 geometric mean: 5.60127371044063"
## [1] "Male first author team size 2018 geometric mean: 4.63867520431303"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7500, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.46432158910349"
## [1] "Male last author team size 2018 geometric mean: 4.90764075114238"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6400, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1      1.018
## LastAuthorFemale  1.026 1      1.013
## UniqueAuthors    1.149 4      1.017
## Year             1.157 16      1.005
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3231 -0.2658  0.0116  0.2543  1.4969
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.66624    0.05072   13.14 < 2e-16 ***
## FirstAuthorFemale1 0.00586    0.01430    0.41  0.68196
## LastAuthorFemale1 0.03265    0.01644    1.99  0.04713 *
## UniqueAuthors2    0.40207    0.04511    8.91 < 2e-16 ***
## UniqueAuthors3    0.47936    0.04518   10.61 < 2e-16 ***
## UniqueAuthors4    0.52508    0.04329   12.13 < 2e-16 ***
## UniqueAuthors5    0.64923    0.04106   15.81 < 2e-16 ***
## Year1997          0.03549    0.04580    0.77  0.43842
## Year1998          0.01116    0.04673    0.24  0.81131
## Year1999         -0.04851    0.04533   -1.07  0.28458
```

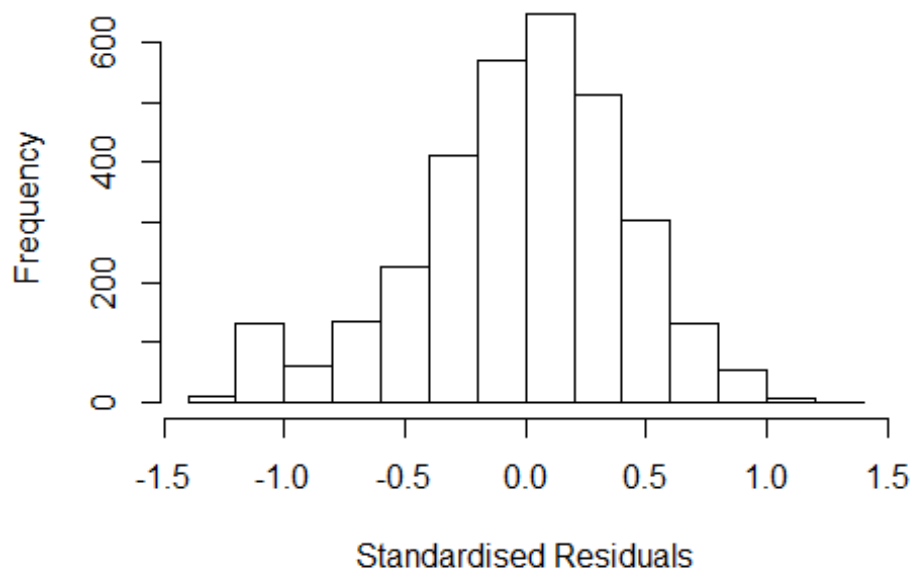
```

## Year2000      0.00759    0.04812    0.16  0.87471
## Year2001     -0.00937    0.04738   -0.20  0.84331
## Year2002     -0.11831    0.04326   -2.73  0.00628 **
## Year2003     -0.14901    0.04611   -3.23  0.00124 **
## Year2004     -0.12452    0.04450   -2.80  0.00517 **
## Year2005     -0.06837    0.04360   -1.57  0.11695
## Year2006     -0.12040    0.04756   -2.53  0.01141 *
## Year2007     -0.16168    0.04499   -3.59  0.00033 ***
## Year2008     -0.09799    0.04514   -2.17  0.03003 *
## Year2009     -0.14464    0.04431   -3.26  0.00111 **
## Year2010     -0.05555    0.04463   -1.24  0.21331
## Year2011     -0.10858    0.04397   -2.47  0.01359 *
## Year2012     -0.10048    0.04447   -2.26  0.02393 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.205, Adjusted R-squared:  0.2
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 272 weights are ~= 1. The remaining 2924 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.861  0.951  0.901  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.13e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1          1.029
## LastAuthorFemale  1.025 1          1.012
## Year              1.076 16          1.002

```



## Residuals from first and last author



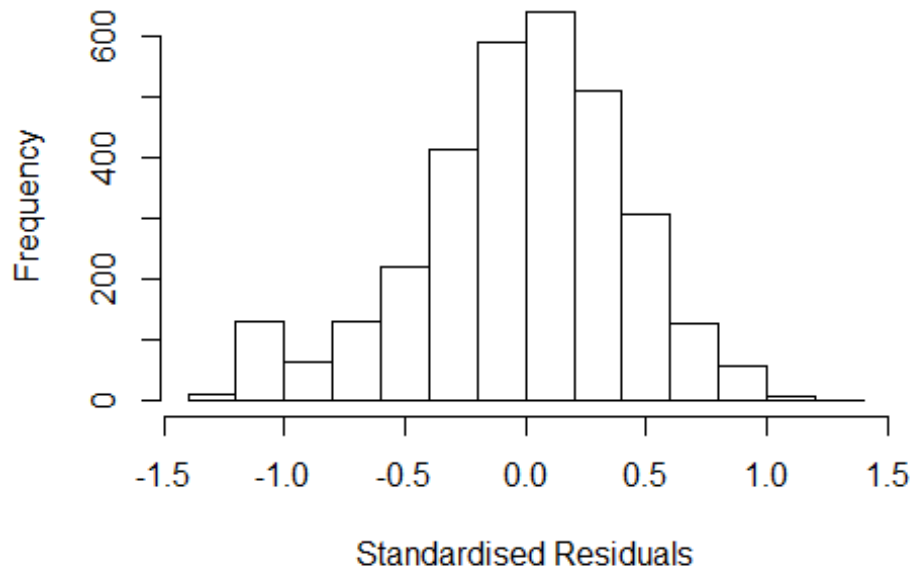
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2936 -0.2774  0.0142  0.2681  1.3081
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12661    0.03291   34.23  <2e-16 ***
## FirstAuthorFemale1 0.04949    0.01565    3.16  0.0016 **
## LastAuthorFemale1 0.03365    0.01694    1.99  0.0471 *
## Year1997         0.05888    0.04467    1.32  0.1875
## Year1998         0.02882    0.04327    0.67  0.5055
## Year1999        -0.00997    0.04418   -0.23  0.8215
## Year2000         0.08388    0.04749    1.77  0.0775 .
## Year2001         0.05695    0.04499    1.27  0.2057
## Year2002        -0.05317    0.04448   -1.20  0.2320
## Year2003        -0.10816    0.04810   -2.25  0.0246 *
## Year2004        -0.07865    0.04616   -1.70  0.0885 .
## Year2005         0.00756    0.04398    0.17  0.8636
```

```

## Year2006      -0.04660    0.04916   -0.95    0.3432
## Year2007      -0.10916    0.04610   -2.37    0.0180 *
## Year2008      -0.05321    0.04627   -1.15    0.2502
## Year2009      -0.08218    0.04432   -1.85    0.0638 .
## Year2010       0.03567    0.04327    0.82    0.4098
## Year2011      -0.04002    0.04681   -0.85    0.3927
## Year2012      -0.01321    0.04542   -0.29    0.7712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0221, Adjusted R-squared:  0.0165
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 266 weights are ~= 1. The remaining 2930 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.859  0.950  0.892  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.054 1      1.027
## Year      1.054 16      1.002

```

## Residuals from first author



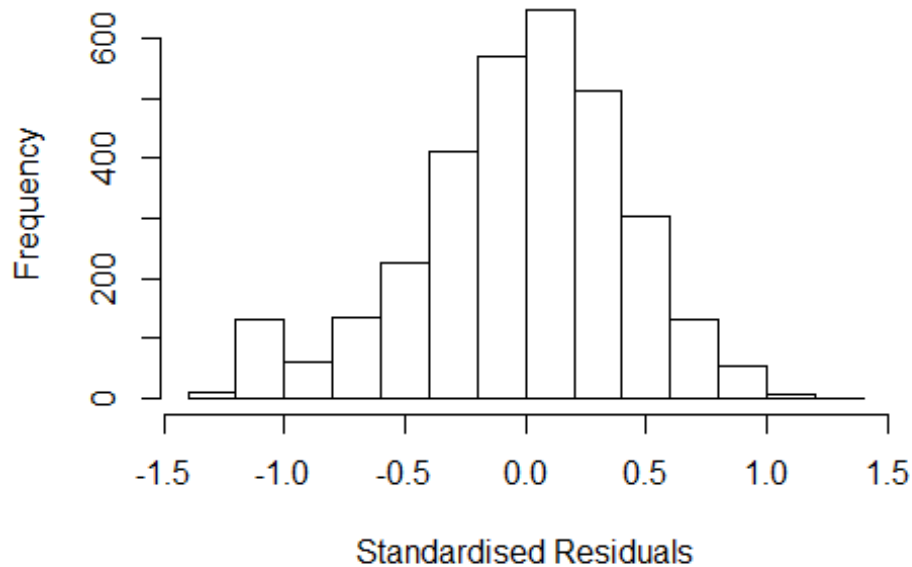
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2691 -0.2785 0.0138 0.2668 1.2988
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13305 0.03252 34.85 <2e-16 ***
## FirstAuthorFemale1 0.05143 0.01575 3.27 0.0011 **
## Year1997 0.06126 0.04447 1.38 0.1684
## Year1998 0.02784 0.04309 0.65 0.5182
## Year1999 -0.00787 0.04406 -0.18 0.8583
## Year2000 0.08463 0.04732 1.79 0.0738 .
## Year2001 0.05913 0.04504 1.31 0.1893
## Year2002 -0.05298 0.04442 -1.19 0.2331
## Year2003 -0.10900 0.04792 -2.27 0.0230 *
## Year2004 -0.07829 0.04604 -1.70 0.0891 .
## Year2005 0.00915 0.04385 0.21 0.8348
## Year2006 -0.04453 0.04904 -0.91 0.3640
```

```

## Year2007          -0.10799    0.04608   -2.34    0.0191 *
## Year2008          -0.05123    0.04622   -1.11    0.2678
## Year2009          -0.08130    0.04445   -1.83    0.0675 .
## Year2010           0.03887    0.04324    0.90    0.3687
## Year2011          -0.03770    0.04664   -0.81    0.4189
## Year2012          -0.00899    0.04526   -0.20    0.8426
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0208, Adjusted R-squared:  0.0155
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 275 weights are ~= 1. The remaining 2921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.270  0.859  0.950  0.892  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.011
## Year              1.021 16      1.001

```

## Residuals from last author



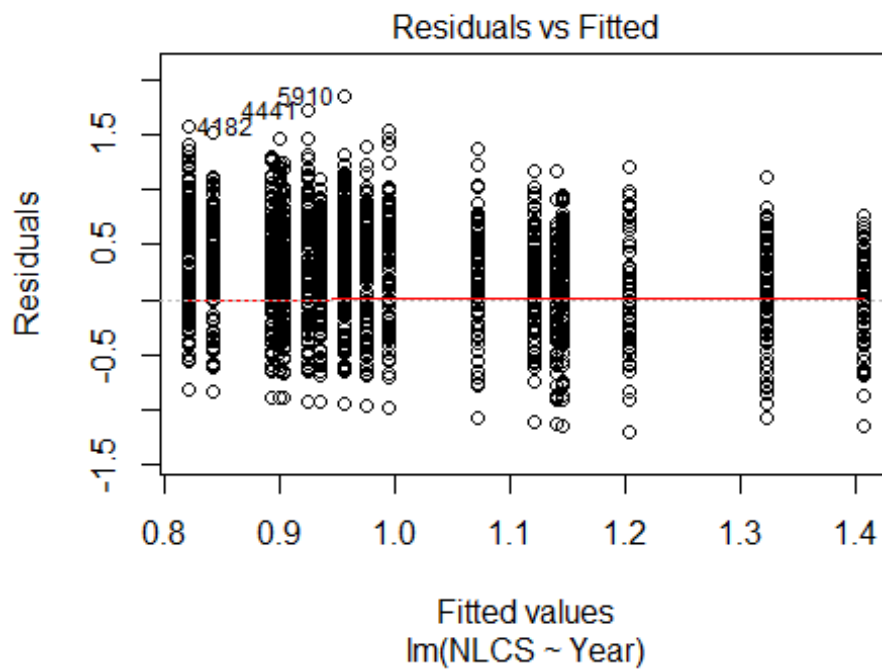
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2698 -0.2728 0.0191 0.2716 1.3312
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14317 0.03259 35.07 <2e-16 ***
## LastAuthorFemale1 0.03734 0.01701 2.20 0.028 *
## Year1997 0.06242 0.04451 1.40 0.161
## Year1998 0.03159 0.04318 0.73 0.464
## Year1999 -0.00730 0.04398 -0.17 0.868
## Year2000 0.08927 0.04677 1.91 0.056 .
## Year2001 0.05939 0.04495 1.32 0.186
## Year2002 -0.04641 0.04426 -1.05 0.294
## Year2003 -0.10149 0.04786 -2.12 0.034 *
## Year2004 -0.07186 0.04612 -1.56 0.119
## Year2005 0.01233 0.04356 0.28 0.777
## Year2006 -0.04335 0.04910 -0.88 0.377
```

```

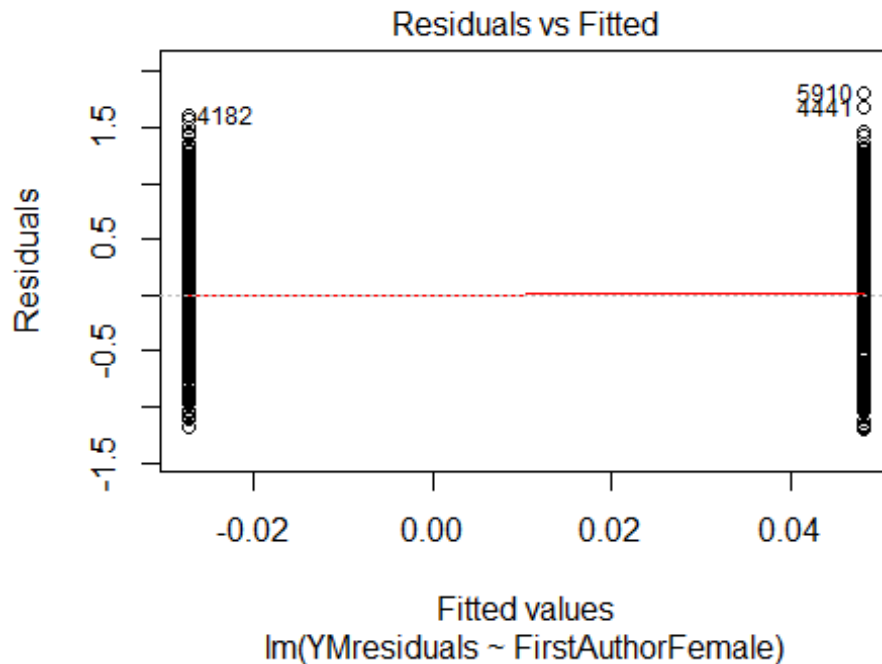
## Year2007          -0.10359      0.04608      -2.25      0.025 *
## Year2008          -0.04483      0.04609      -0.97      0.331
## Year2009          -0.07241      0.04408      -1.64      0.101
## Year2010           0.04328      0.04310       1.00      0.315
## Year2011          -0.03332      0.04670      -0.71      0.476
## Year2012          -0.00558      0.04517      -0.12      0.902
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.0185, Adjusted R-squared:  0.0133
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 290 weights are ~= 1. The remaining 2906 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.246  0.861  0.949  0.891  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.13e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3196"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2724"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  224  231  215  222  207  272  255  271  269  291  322  350  346  293  315
## 2011 2012
##  304  343
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   87   59   67   89  100   91  116  153  146  168  170  178  201  163  144
## 2011 2012

```

```
## 143 165
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 82 54 63 64 78 70 99 112 129 122 145 153 175 128 123
## 2011 2012
## 122 134
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 58, df = 16, p-value = 1e-06
```



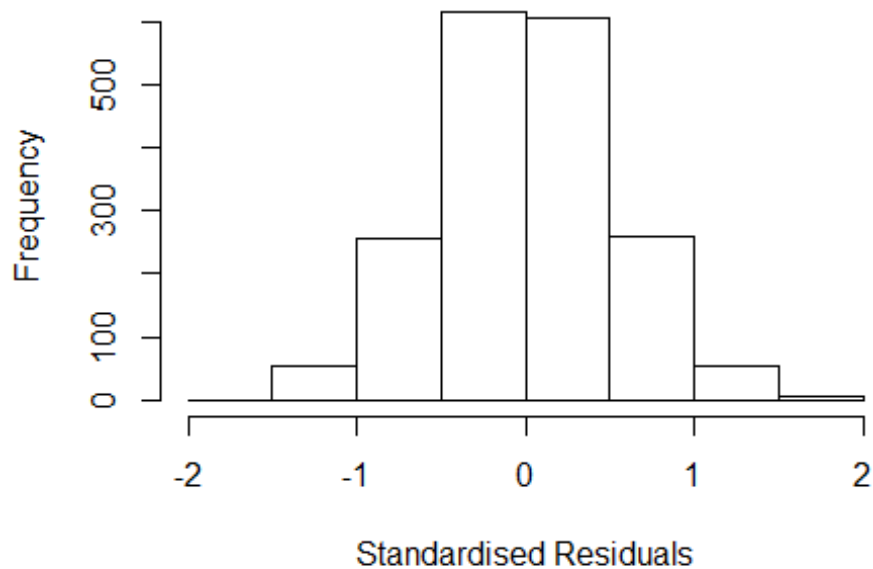
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 5.72469956060036"
## [1] "Male first author team size 2018 geometric mean: 4.76066878002908"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.1221212924666"
## [1] "Male last author team size 2018 geometric mean: 5.15810060505026"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1100, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.165 1      1.079
## LastAuthorFemale  1.140 1      1.068
## UniqueAuthors     1.234 4      1.027
## Year               1.299 16     1.008
```



## Residuals from first and last author and team size



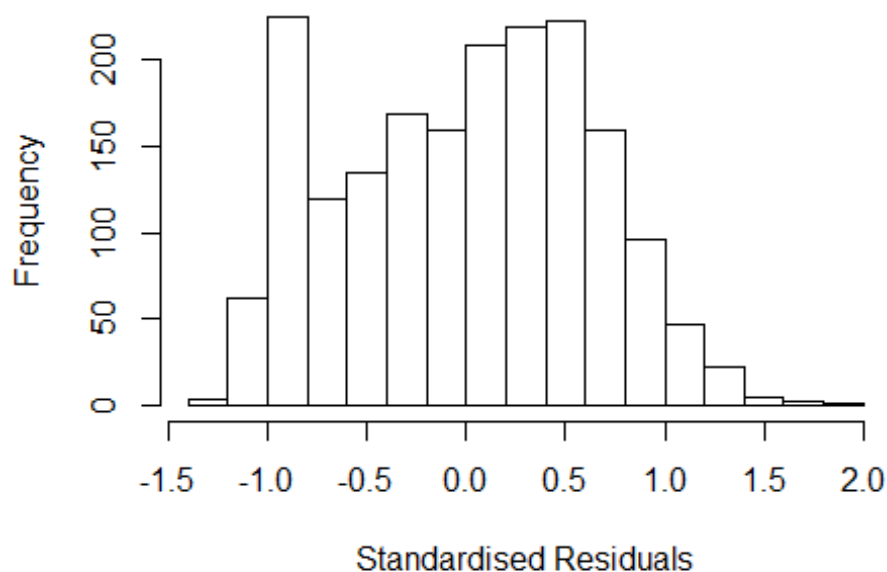
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.507663 -0.361075 0.000356 0.369940 1.861155
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5485 0.0799 6.86 9.1e-12 ***
## FirstAuthorFemale1 0.0412 0.0288 1.43 0.1531
## LastAuthorFemale1 -0.0318 0.0335 -0.95 0.3424
## UniqueAuthors2 0.2105 0.0467 4.51 7.0e-06 ***
## UniqueAuthors3 0.3618 0.0499 7.25 6.1e-13 ***
## UniqueAuthors4 0.6903 0.0477 14.47 < 2e-16 ***
## UniqueAuthors5 0.9026 0.0340 26.51 < 2e-16 ***
## Year1997 0.0884 0.1106 0.80 0.4241
## Year1998 0.2638 0.0888 2.97 0.0030 **
## Year1999 0.1833 0.0973 1.88 0.0597 .
```

```

## Year2000          0.0411      0.0919      0.45      0.6549
## Year2001          0.0107      0.0988      0.11      0.9135
## Year2002         -0.0247      0.0936     -0.26      0.7915
## Year2003         -0.2037      0.0874     -2.33      0.0199 *
## Year2004         -0.2115      0.0876     -2.41      0.0159 *
## Year2005         -0.1619      0.0868     -1.86      0.0623 .
## Year2006         -0.1662      0.0888     -1.87      0.0613 .
## Year2007         -0.1749      0.0881     -1.98      0.0473 *
## Year2008         -0.2594      0.0883     -2.94      0.0034 **
## Year2009         -0.1642      0.0932     -1.76      0.0783 .
## Year2010         -0.1592      0.0956     -1.67      0.0961 .
## Year2011         -0.1087      0.0922     -1.18      0.2384
## Year2012         -0.1960      0.0924     -2.12      0.0339 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.512
## Multiple R-squared:  0.345, Adjusted R-squared:  0.337
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1695 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.158  0.867  0.949  0.900  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.40e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.040 1      1.020
## LastAuthorFemale  1.040 1      1.020
## Year              1.058 16      1.002

```

## Residuals from first and last author



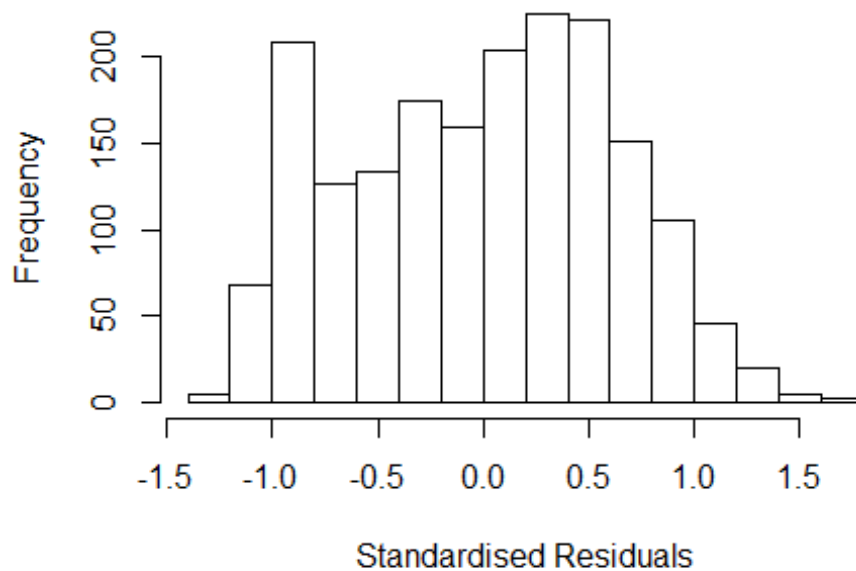
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2918 -0.5244 0.0467 0.4748 1.8286
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0627 0.0704 15.09 < 2e-16 ***
## FirstAuthorFemale1 0.1124 0.0334 3.36 0.00078 ***
## LastAuthorFemale1 -0.1078 0.0413 -2.61 0.00914 **
## Year1997 0.1111 0.1069 1.04 0.29904
## Year1998 0.3196 0.0863 3.70 0.00022 ***
## Year1999 0.2720 0.0994 2.74 0.00628 **
## Year2000 0.1167 0.0922 1.27 0.20553
## Year2001 0.0602 0.0964 0.62 0.53246
## Year2002 0.0696 0.0965 0.72 0.47080
## Year2003 -0.2217 0.0922 -2.40 0.01628 *
## Year2004 -0.2328 0.0922 -2.53 0.01162 *
## Year2005 -0.1715 0.0920 -1.86 0.06249 .
```

```

## Year2006          -0.1563      0.0927   -1.69  0.09200 .
## Year2007          -0.1877      0.0904   -2.08  0.03802 *
## Year2008          -0.2485      0.0909   -2.73  0.00635 **
## Year2009          -0.1096      0.0992   -1.11  0.26917
## Year2010          -0.0678      0.0966   -0.70  0.48278
## Year2011          -0.0468      0.0912   -0.51  0.60771
## Year2012          -0.0869      0.0951   -0.91  0.36108
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0599, Adjusted R-squared:  0.0506
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 134 weights are ~= 1. The remaining 1719 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.459  0.875   0.946   0.924   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.40e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling            cov
##      "bisquare"      "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1      1.014
## Year              1.028 16      1.001

```

## Residuals from first author



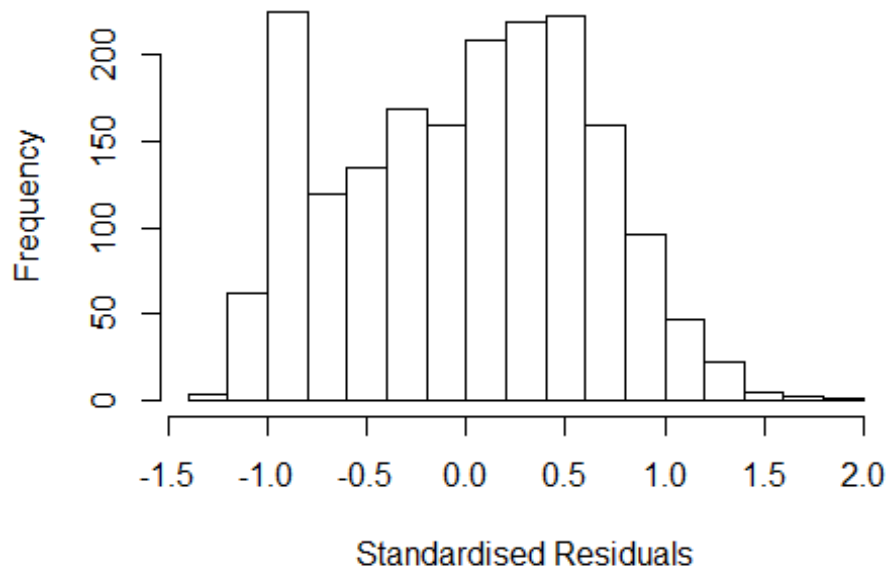
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2630 -0.5263 0.0496 0.4824 1.7588
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0518 0.0708 14.86 < 2e-16 ***
## FirstAuthorFemale1 0.0978 0.0337 2.90 0.00372 **
## Year1997 0.1057 0.1073 0.98 0.32497
## Year1998 0.3251 0.0867 3.75 0.00018 ***
## Year1999 0.2767 0.0994 2.78 0.00543 **
## Year2000 0.1135 0.0929 1.22 0.22232
## Year2001 0.0575 0.0969 0.59 0.55284
## Year2002 0.0664 0.0963 0.69 0.49068
## Year2003 -0.2177 0.0925 -2.35 0.01873 *
## Year2004 -0.2367 0.0931 -2.54 0.01105 *
## Year2005 -0.1760 0.0928 -1.90 0.05803 .
## Year2006 -0.1635 0.0933 -1.75 0.07972 .
```

```

## Year2007          -0.1899      0.0913   -2.08  0.03771 *
## Year2008          -0.2555      0.0914   -2.80  0.00523 **
## Year2009          -0.1142      0.0998   -1.14  0.25277
## Year2010          -0.0745      0.0969   -0.77  0.44217
## Year2011          -0.0514      0.0916   -0.56  0.57480
## Year2012          -0.0994      0.0953   -1.04  0.29731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.687
## Multiple R-squared:  0.0562, Adjusted R-squared:  0.0475
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1722 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.492  0.876  0.945  0.924  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.40e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.014
## Year            1.029 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2217 -0.5249 0.0528 0.4820 1.8715
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0809 0.0711 15.20 < 2e-16 ***
## LastAuthorFemale1 -0.0842 0.0405 -2.08 0.03778 *
## Year1997 0.1198 0.1072 1.12 0.26406
## Year1998 0.3278 0.0871 3.77 0.00017 ***
## Year1999 0.3007 0.0990 3.04 0.00243 **
## Year2000 0.1408 0.0921 1.53 0.12647
## Year2001 0.0807 0.0978 0.83 0.40919
## Year2002 0.0798 0.0964 0.83 0.40779
## Year2003 -0.2134 0.0936 -2.28 0.02276 *
## Year2004 -0.2218 0.0929 -2.39 0.01704 *
## Year2005 -0.1584 0.0926 -1.71 0.08739 .
## Year2006 -0.1451 0.0931 -1.56 0.11942
```

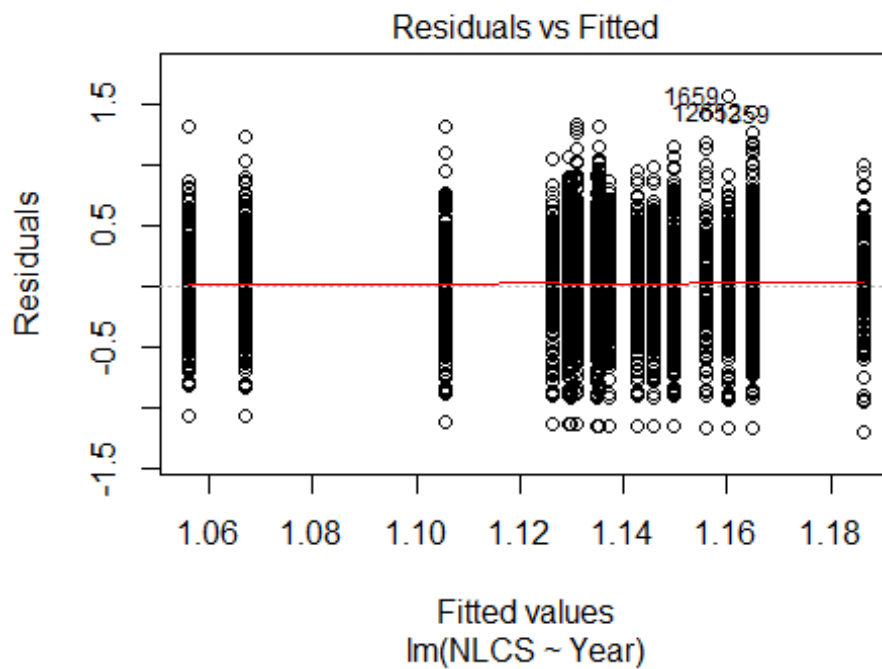
```

## Year2007          -0.1699      0.0911    -1.87   0.06234 .
## Year2008          -0.2288      0.0910    -2.52   0.01198 *
## Year2009          -0.0877      0.0997    -0.88   0.37893
## Year2010          -0.0464      0.0976    -0.48   0.63433
## Year2011          -0.0312      0.0918    -0.34   0.73422
## Year2012          -0.0592      0.0952    -0.62   0.53370
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.692
## Multiple R-squared:  0.0538, Adjusted R-squared:  0.045
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.445  0.874  0.946  0.924  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.40e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1853"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2725"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  537  548  521  556  593  630  589  483  450  493  547  576  591  659  621
## 2011 2012
##  690  661
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  148  143  173  195  186  153  286  254  252  279  290  311  336  379  315
## 2011 2012

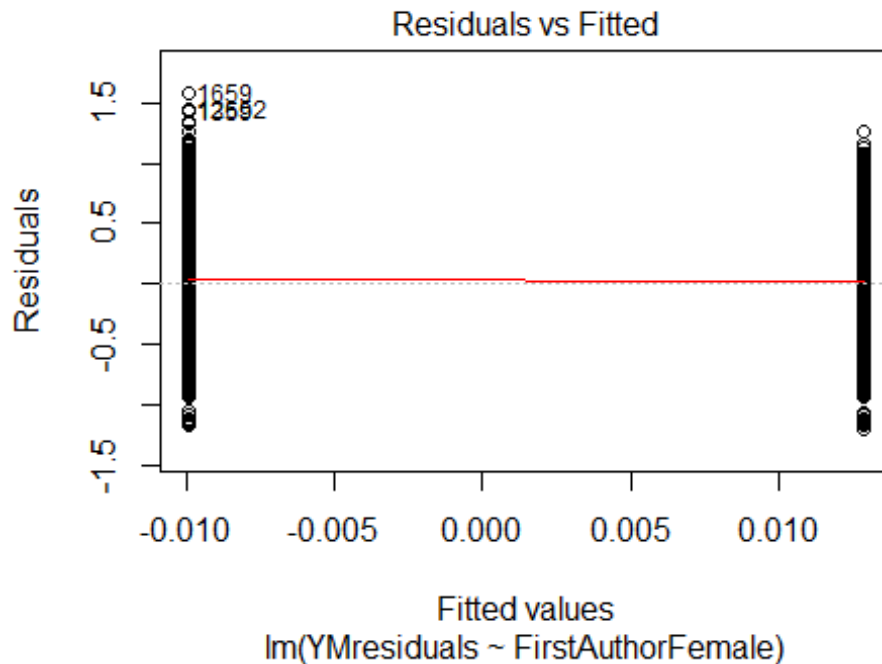
```



```
## 399 363
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 122 154 161 169 136 230 188 201 211 236 254 268 301 275
## 2011 2012
## 354 318
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 67, df = 16, p-value = 3e-08
```

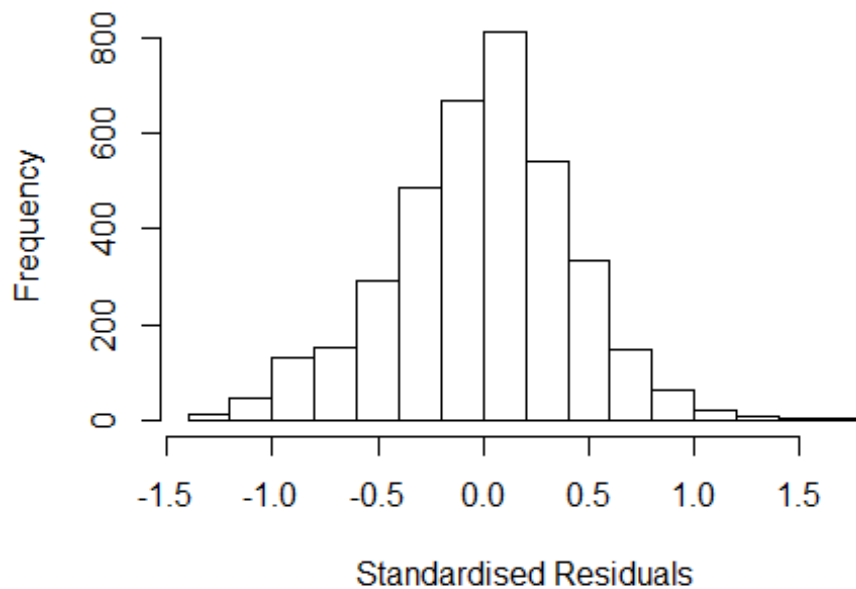


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.2, df = 1, p-value = 0.007
```



```
## [1] "Female first author team size 2018 geometric mean: 4.838011093562"
## [1] "Male first author team size 2018 geometric mean: 4.5425209556933"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.55008363343574"
## [1] "Male last author team size 2018 geometric mean: 4.79443014554387"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1          1.019
## LastAuthorFemale  1.026 1          1.013
## UniqueAuthors    1.127 4          1.015
## Year             1.152 16          1.004
```

## Residuals from first and last author and team size



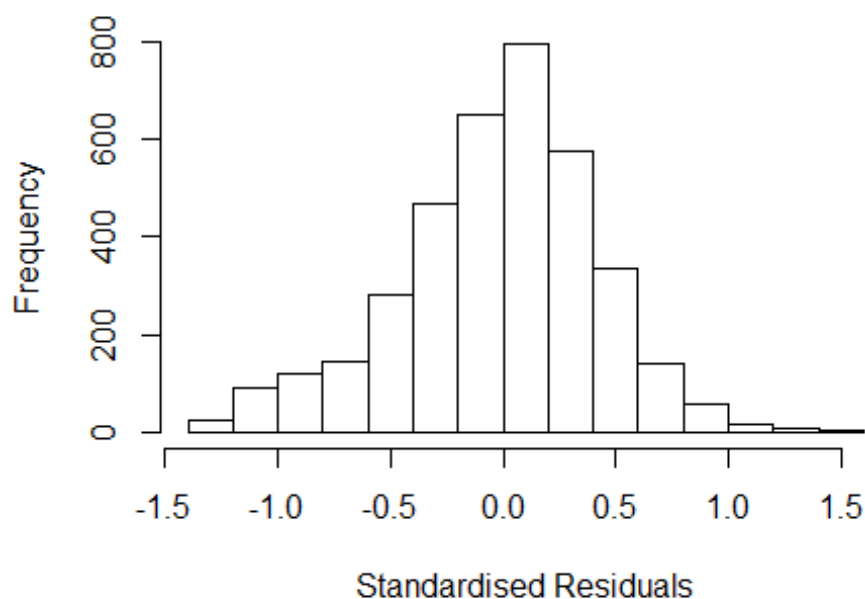
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2872 -0.2673 0.0178 0.2642 1.7710
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0260 0.0503 20.38 < 2e-16 ***
## FirstAuthorFemale1 -0.0102 0.0142 -0.71 0.47566
## LastAuthorFemale1 -0.0037 0.0166 -0.22 0.82358
## UniqueAuthors2 0.1404 0.0376 3.73 0.00019 ***
## UniqueAuthors3 0.2137 0.0343 6.22 5.4e-10 ***
## UniqueAuthors4 0.2450 0.0342 7.17 8.9e-13 ***
## UniqueAuthors5 0.3092 0.0307 10.07 < 2e-16 ***
## Year1997 -0.0539 0.0618 -0.87 0.38303
## Year1998 -0.0670 0.0564 -1.19 0.23524
## Year1999 -0.0940 0.0532 -1.77 0.07748 .
```

```

## Year2000          -0.0808      0.0520   -1.55   0.12006
## Year2001          -0.0817      0.0574   -1.42   0.15502
## Year2002          -0.0894      0.0496   -1.80   0.07165 .
## Year2003          -0.1108      0.0498   -2.22   0.02618 *
## Year2004          -0.0966      0.0500   -1.93   0.05340 .
## Year2005          -0.1592      0.0508   -3.14   0.00172 **
## Year2006          -0.1580      0.0497   -3.18   0.00148 **
## Year2007          -0.1627      0.0490   -3.32   0.00090 ***
## Year2008          -0.0759      0.0491   -1.54   0.12251
## Year2009          -0.0789      0.0495   -1.59   0.11108
## Year2010          -0.0798      0.0497   -1.61   0.10832
## Year2011          -0.0480      0.0505   -0.95   0.34154
## Year2012          -0.1040      0.0509   -2.04   0.04104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0641, Adjusted R-squared:  0.0585
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 313 weights are ~= 1. The remaining 3395 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0073 0.8590 0.9500 0.8930 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.70e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.030 1          1.015
## LastAuthorFemale 1.025 1          1.012
## Year          1.040 16          1.001

```

## Residuals from first and last author



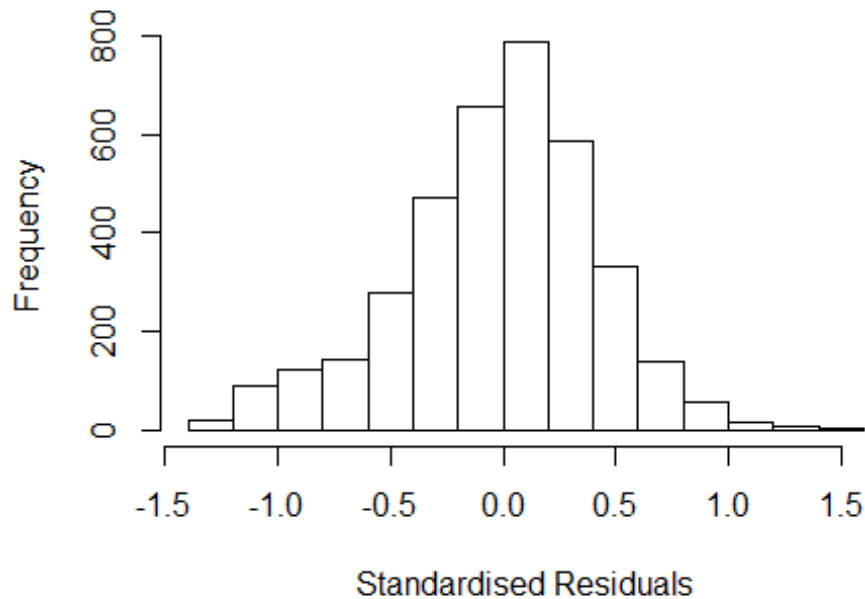
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2352 -0.2744  0.0179  0.2659  1.5598
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2219    0.0420   29.06  <2e-16 ***
## FirstAuthorFemale1  0.0133    0.0145    0.92  0.3592
## LastAuthorFemale1 -0.0128    0.0172   -0.74  0.4579
## Year1997          -0.0524    0.0623   -0.84  0.4010
## Year1998          -0.0517    0.0566   -0.91  0.3611
## Year1999          -0.0728    0.0529   -1.38  0.1685
## Year2000          -0.0547    0.0523   -1.05  0.2957
## Year2001          -0.0617    0.0572   -1.08  0.2802
## Year2002          -0.0475    0.0496   -0.96  0.3388
## Year2003          -0.0803    0.0494   -1.63  0.1038
## Year2004          -0.0685    0.0495   -1.38  0.1662
## Year2005          -0.1185    0.0505   -2.34  0.0191 *
```

```

## Year2006          -0.1287      0.0495   -2.60   0.0094 **
## Year2007          -0.1505      0.0486   -3.09   0.0020 **
## Year2008          -0.0434      0.0484   -0.90   0.3700
## Year2009          -0.0614      0.0490   -1.25   0.2098
## Year2010          -0.0320      0.0494   -0.65   0.5167
## Year2011          -0.0107      0.0501   -0.21   0.8315
## Year2012          -0.0579      0.0505   -1.14   0.2524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.00886,    Adjusted R-squared:  0.00402
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 309 weights are ~= 1. The remaining 3399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0952 0.8600 0.9500 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## Year              1.022 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2314 -0.2755  0.0176  0.2644  1.5618
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2198     0.0421   28.96  <2e-16 ***
## FirstAuthorFemale1  0.0116     0.0144    0.80   0.4222
## Year1997         -0.0530     0.0624   -0.85   0.3960
## Year1998         -0.0516     0.0566   -0.91   0.3624
## Year1999         -0.0736     0.0530   -1.39   0.1648
## Year2000         -0.0554     0.0524   -1.06   0.2910
## Year2001         -0.0613     0.0572   -1.07   0.2838
## Year2002         -0.0469     0.0497   -0.94   0.3454
## Year2003         -0.0804     0.0495   -1.62   0.1046
## Year2004         -0.0687     0.0496   -1.39   0.1656
## Year2005         -0.1182     0.0506   -2.33   0.0196 *
## Year2006         -0.1290     0.0496   -2.60   0.0093 **
```

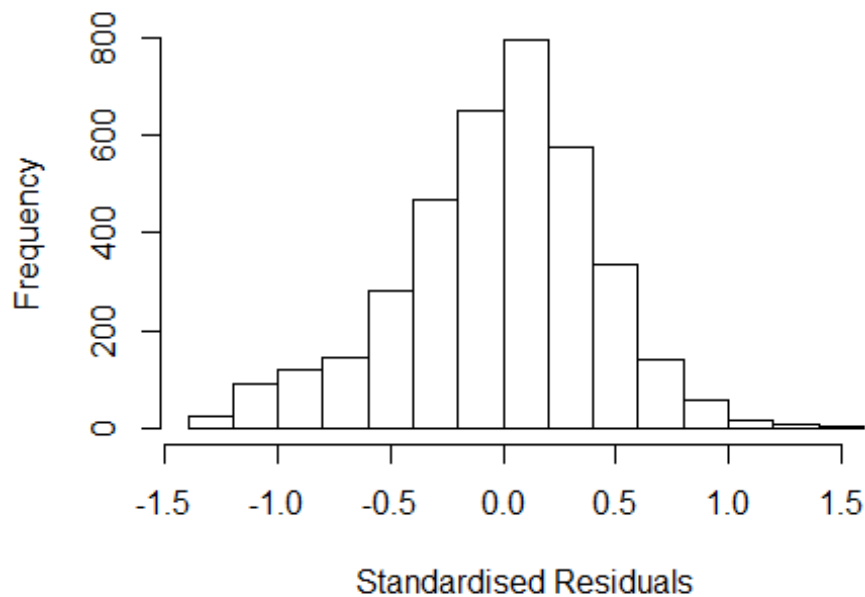
```

## Year2007          -0.1512      0.0487   -3.11   0.0019 **
## Year2008          -0.0443      0.0484   -0.92   0.3601
## Year2009          -0.0615      0.0491   -1.25   0.2101
## Year2010          -0.0328      0.0494   -0.66   0.5064
## Year2011          -0.0109      0.0501   -0.22   0.8280
## Year2012          -0.0591      0.0506   -1.17   0.2428
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.4
## Multiple R-squared:  0.00873,    Adjusted R-squared:  0.00416
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 317 weights are ~= 1. The remaining 3391 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0939 0.8600 0.9500 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1          1.008
## Year            1.017 16          1.001

```



## Residuals from last author



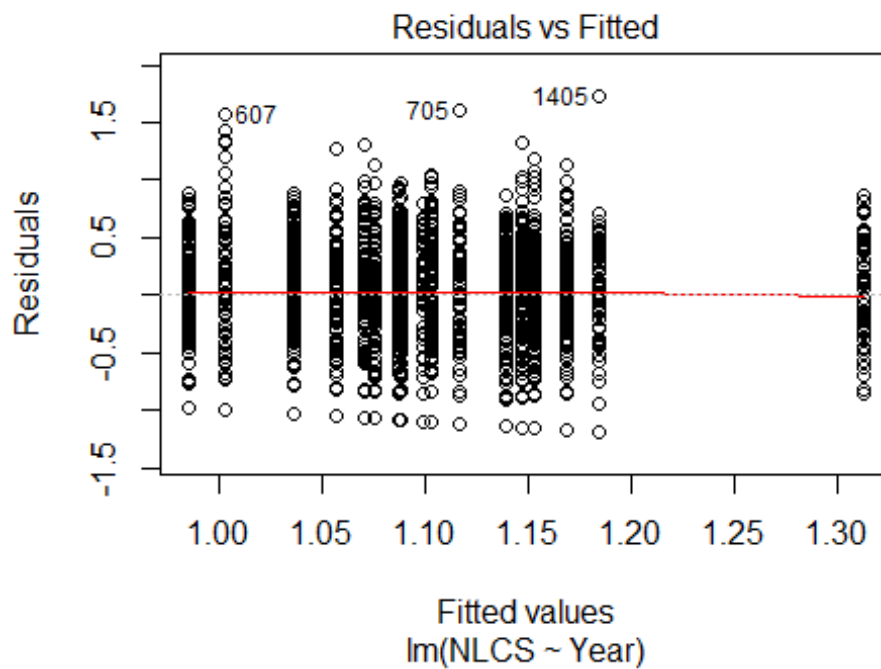
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2259 -0.2762  0.0194  0.2642  1.5551
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22589    0.04188   29.27  <2e-16 ***
## LastAuthorFemale1 -0.01053    0.01717   -0.61  0.5398
## Year1997        -0.05194    0.06236   -0.83  0.4049
## Year1998        -0.05098    0.05659   -0.90  0.3677
## Year1999        -0.07266    0.05294   -1.37  0.1700
## Year2000        -0.05409    0.05242   -1.03  0.3022
## Year2001        -0.06005    0.05718   -1.05  0.2937
## Year2002        -0.04597    0.04969   -0.93  0.3549
## Year2003        -0.07884    0.04948   -1.59  0.1111
## Year2004        -0.06751    0.04960   -1.36  0.1735
## Year2005        -0.11631    0.05050   -2.30  0.0213 *
## Year2006        -0.12764    0.04962   -2.57  0.0101 *
```

```

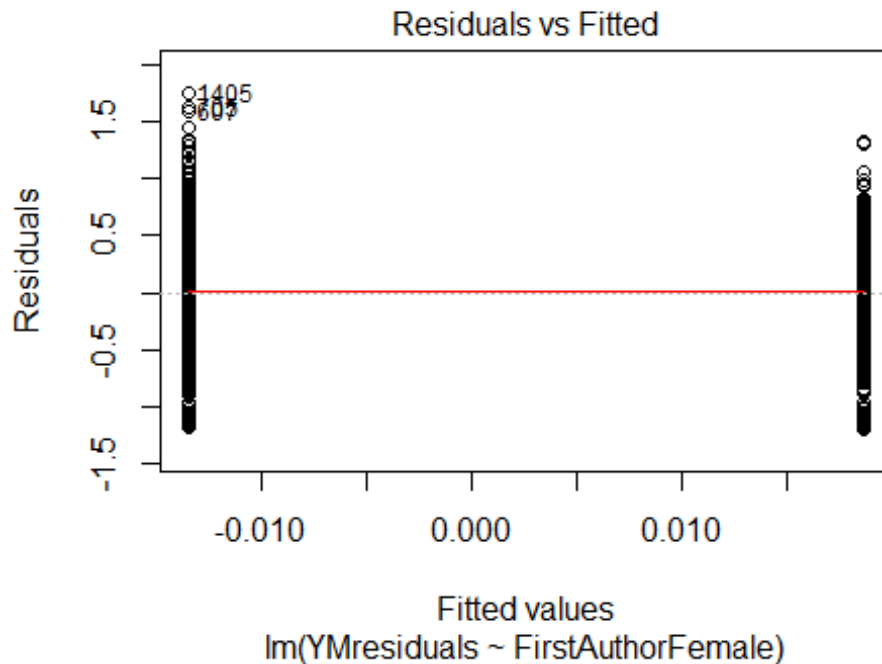
## Year2007          -0.14917      0.04875    -3.06    0.0022 **
## Year2008          -0.04213      0.04843    -0.87    0.3844
## Year2009          -0.05973      0.04900    -1.22    0.2229
## Year2010          -0.03048      0.04944    -0.62    0.5376
## Year2011          -0.00859      0.05005    -0.17    0.8637
## Year2012          -0.05637      0.05066    -1.11    0.2658
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.401
## Multiple R-squared:  0.00863,    Adjusted R-squared:  0.00406
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 311 weights are ~= 1. The remaining 3397 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0984 0.8610 0.9500 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3708"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2726"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 262 289 268 303 309 336 271 272 223 251 237 255 240 225 248
## 2011 2012
## 268 239
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 57 79 66 110 70 49 120 152 127 135 109 131 149 141 136
## 2011 2012

```

```
## 147 129
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 66 60 93 59 43 97 107 98 100 88 113 125 119 114
## 2011 2012
## 127 108
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```

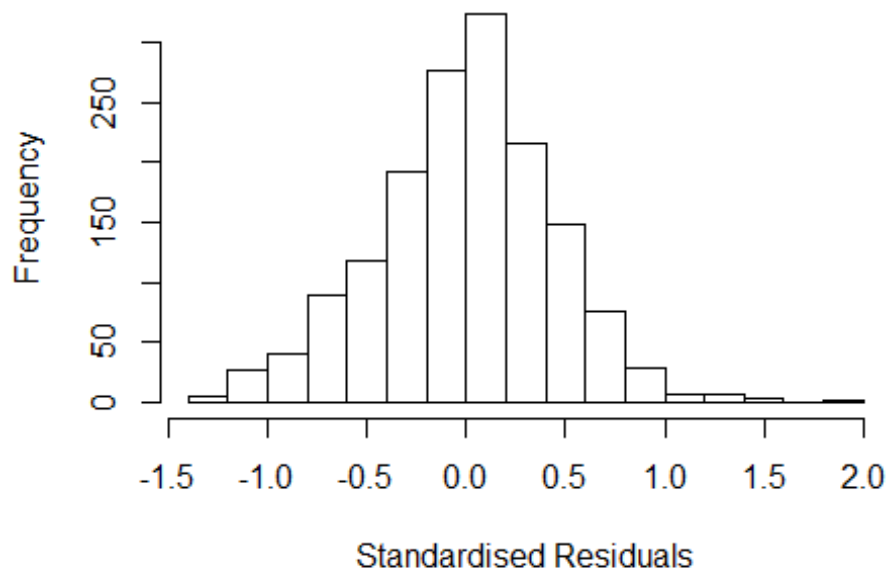


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 15, df = 1, p-value = 1e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 5.18417780327572"
## [1] "Male first author team size 2018 geometric mean: 4.76224967678331"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.97241045201924"
## [1] "Male last author team size 2018 geometric mean: 4.5027750607154"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.089 1      1.043
## LastAuthorFemale  1.098 1      1.048
## UniqueAuthors    1.272 4      1.031
## Year              1.350 16     1.009
```

## Residuals from first and last author and team size



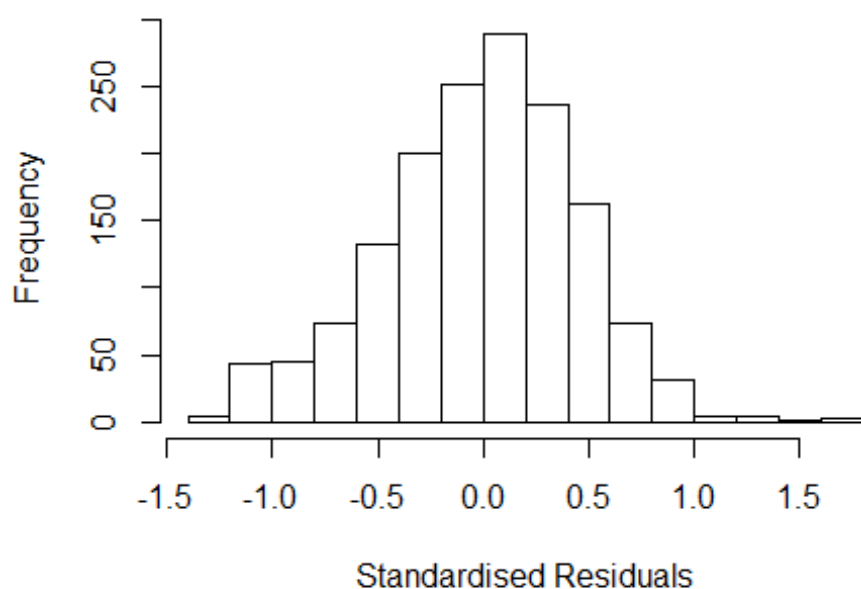
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2828 -0.2825 0.0169 0.2802 1.8535
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20965 0.08203 14.75 < 2e-16 ***
## FirstAuthorFemale1 -0.00435 0.02342 -0.19 0.85264
## LastAuthorFemale1 -0.01846 0.02786 -0.66 0.50764
## UniqueAuthors2 0.11317 0.05428 2.08 0.03726 *
## UniqueAuthors3 0.15093 0.05113 2.95 0.00320 **
## UniqueAuthors4 0.26173 0.05220 5.01 6.0e-07 ***
## UniqueAuthors5 0.35393 0.04433 7.98 2.8e-15 ***
## Year1997 -0.48513 0.11808 -4.11 4.2e-05 ***
## Year1998 -0.29434 0.09953 -2.96 0.00315 **
## Year1999 -0.37925 0.08825 -4.30 1.8e-05 ***
```

```

## Year2000      -0.21083    0.09810    -2.15    0.03177 *
## Year2001      -0.23113    0.11204    -2.06    0.03929 *
## Year2002      -0.42327    0.08905    -4.75    2.2e-06 ***
## Year2003      -0.32010    0.08299    -3.86    0.00012 ***
## Year2004      -0.25199    0.08656    -2.91    0.00365 **
## Year2005      -0.33753    0.08443    -4.00    6.7e-05 ***
## Year2006      -0.37543    0.08749    -4.29    1.9e-05 ***
## Year2007      -0.32071    0.08542    -3.75    0.00018 ***
## Year2008      -0.33032    0.08206    -4.03    6.0e-05 ***
## Year2009      -0.23770    0.08313    -2.86    0.00430 **
## Year2010      -0.29933    0.08607    -3.48    0.00052 ***
## Year2011      -0.28075    0.08413    -3.34    0.00087 ***
## Year2012      -0.25842    0.08569    -3.02    0.00260 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.421
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0885
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1428 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0139 0.8570 0.9510 0.8970 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.41e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1          1.028
## LastAuthorFemale 1.070 1          1.034
## Year 1.104 16          1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2318 -0.3017 0.0177 0.2976 1.7152
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3442 0.0807 16.65 < 2e-16 ***
## FirstAuthorFemale1 0.0257 0.0241 1.07 0.28675
## LastAuthorFemale1 -0.0156 0.0299 -0.52 0.60107
## Year1997 -0.4311 0.1229 -3.51 0.00046 ***
## Year1998 -0.2667 0.1063 -2.51 0.01224 *
## Year1999 -0.3149 0.0932 -3.38 0.00075 ***
## Year2000 -0.1474 0.1010 -1.46 0.14461
## Year2001 -0.2196 0.1168 -1.88 0.06019 .
## Year2002 -0.3243 0.0976 -3.32 0.00091 ***
## Year2003 -0.2305 0.0882 -2.61 0.00904 **
## Year2004 -0.1555 0.0916 -1.70 0.08998 .
## Year2005 -0.2194 0.0910 -2.41 0.01602 *
```

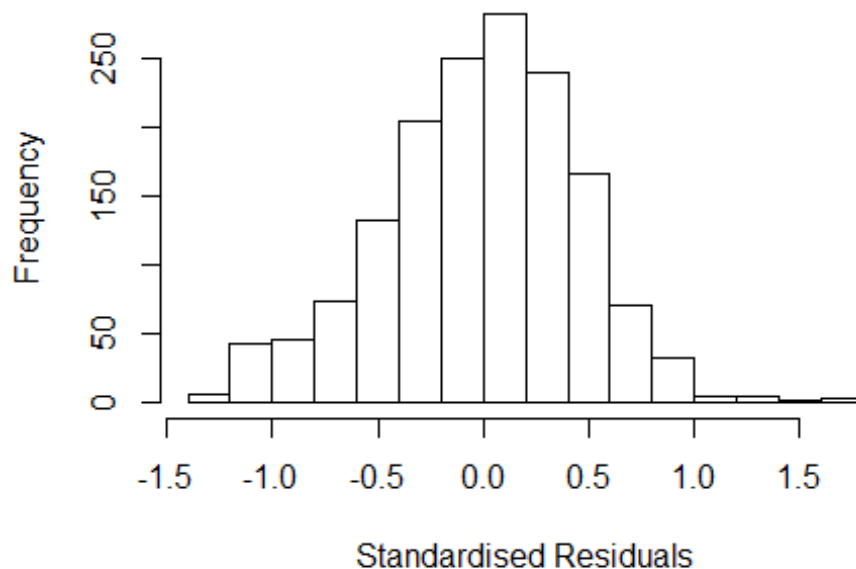
```

## Year2006          -0.2564      0.0940    -2.73  0.00643 **
## Year2007          -0.2345      0.0918    -2.56  0.01068 *
## Year2008          -0.2408      0.0882    -2.73  0.00638 **
## Year2009          -0.1480      0.0873    -1.70  0.09026 .
## Year2010          -0.1829      0.0920    -1.99  0.04693 *
## Year2011          -0.1485      0.0885    -1.68  0.09356 .
## Year2012          -0.1225      0.0908    -1.35  0.17739
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.0315, Adjusted R-squared:  0.0202
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1434 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.101  0.872  0.951  0.900  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.41e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.045 1          1.022
## Year              1.045 16          1.001

```



## Residuals from first author



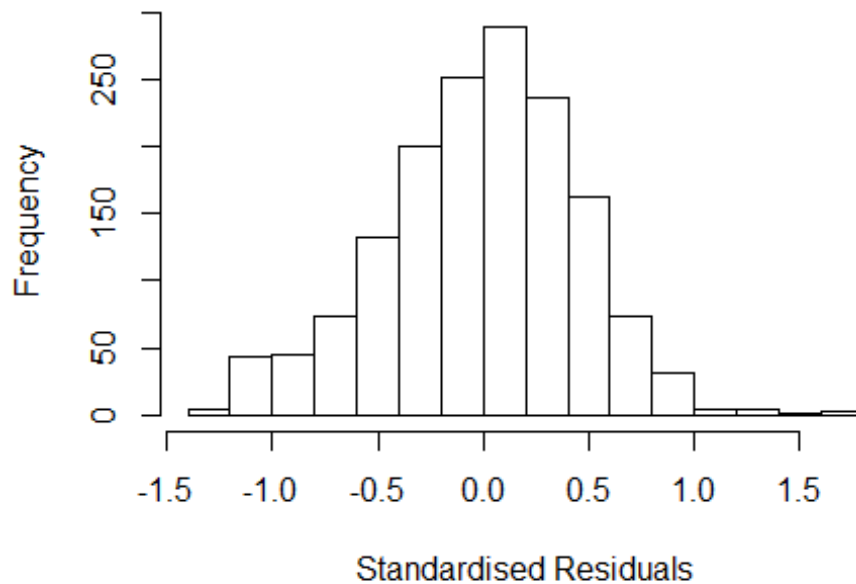
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2425 -0.2988 0.0167 0.3014 1.7197
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3398 0.0803 16.69 < 2e-16 ***
## FirstAuthorFemale1 0.0243 0.0240 1.01 0.31211
## Year1997 -0.4290 0.1228 -3.49 0.00049 ***
## Year1998 -0.2662 0.1065 -2.50 0.01258 *
## Year1999 -0.3139 0.0932 -3.37 0.00078 ***
## Year2000 -0.1474 0.1011 -1.46 0.14505
## Year2001 -0.2179 0.1166 -1.87 0.06187 .
## Year2002 -0.3213 0.0971 -3.31 0.00096 ***
## Year2003 -0.2283 0.0880 -2.59 0.00960 **
## Year2004 -0.1537 0.0916 -1.68 0.09344 .
## Year2005 -0.2167 0.0909 -2.39 0.01719 *
## Year2006 -0.2539 0.0938 -2.71 0.00688 **
```

```

## Year2007          -0.2347      0.0917   -2.56  0.01060 *
## Year2008          -0.2390      0.0881   -2.71  0.00673 **
## Year2009          -0.1463      0.0873   -1.68  0.09373 .
## Year2010          -0.1823      0.0920   -1.98  0.04760 *
## Year2011          -0.1464      0.0884   -1.66  0.09768 .
## Year2012          -0.1215      0.0907   -1.34  0.18050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.0314, Adjusted R-squared:  0.0207
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1434 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0979 0.8720 0.9510 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.41e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.058 1          1.029
## Year            1.058 16          1.002

```

## Residuals from last author



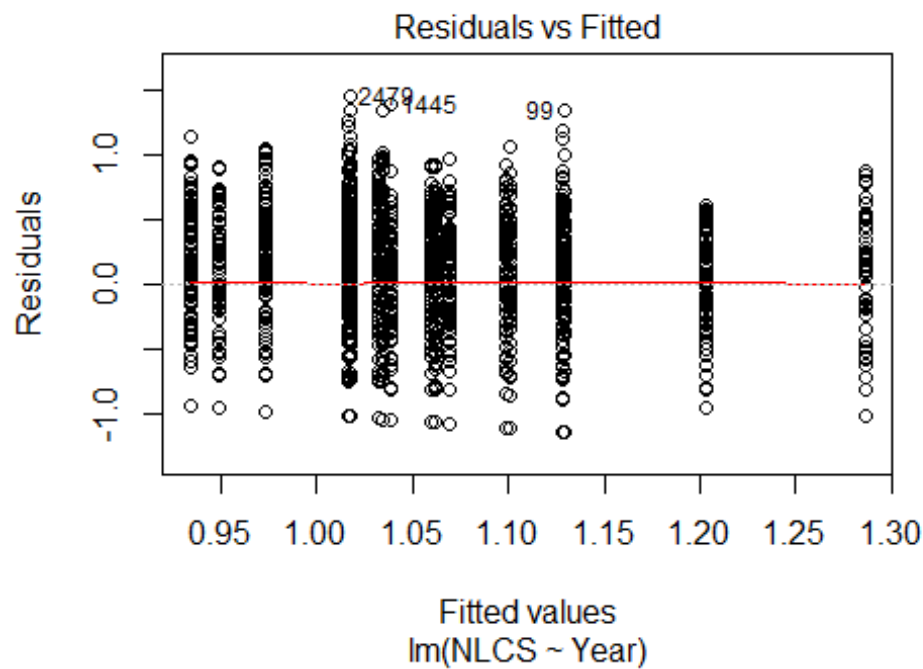
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2343 -0.2995 0.0199 0.3015 1.7043
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3529 0.0799 16.93 < 2e-16 ***
## LastAuthorFemale1 -0.0123 0.0297 -0.41 0.67939
## Year1997 -0.4320 0.1233 -3.50 0.00047 ***
## Year1998 -0.2655 0.1066 -2.49 0.01288 *
## Year1999 -0.3158 0.0930 -3.39 0.00071 ***
## Year2000 -0.1452 0.1007 -1.44 0.14942
## Year2001 -0.2200 0.1163 -1.89 0.05867 .
## Year2002 -0.3242 0.0974 -3.33 0.00089 ***
## Year2003 -0.2289 0.0879 -2.61 0.00927 **
## Year2004 -0.1532 0.0916 -1.67 0.09450 .
## Year2005 -0.2166 0.0908 -2.39 0.01714 *
## Year2006 -0.2546 0.0939 -2.71 0.00677 **
```

```

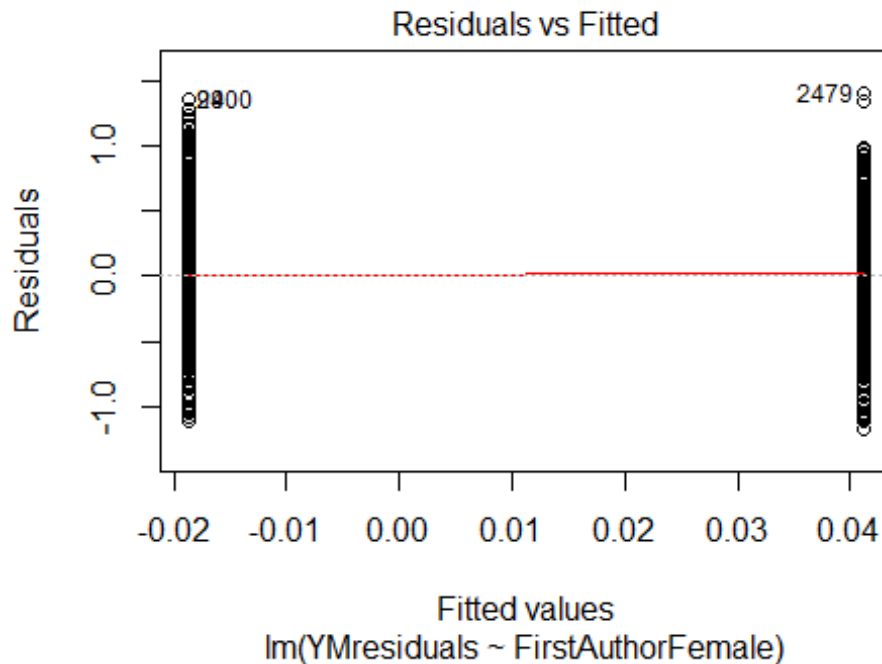
## Year2007          -0.2319      0.0914    -2.54  0.01129 *
## Year2008          -0.2397      0.0880    -2.72  0.00652 **
## Year2009          -0.1471      0.0870    -1.69  0.09110 .
## Year2010          -0.1808      0.0917    -1.97  0.04890 *
## Year2011          -0.1428      0.0880    -1.62  0.10475
## Year2012          -0.1186      0.0904    -1.31  0.18955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.443
## Multiple R-squared:  0.031, Adjusted R-squared:  0.0203
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~ = 1. The remaining 1429 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.106  0.870  0.951  0.900  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.41e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1559"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2727"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 180 168 153 162 150 216 185 122 142 114 141 176 179 196 152
## 2011 2012
## 227 185
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 50 56 84 109 78 105 118 85 105 83 94 132 136 151 116
## 2011 2012

```

```
## 195 149
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 42 50 68 93 70 92 103 74 96 72 79 115 116 133 104
## 2011 2012
## 177 133
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 34, df = 16, p-value = 0.005
```

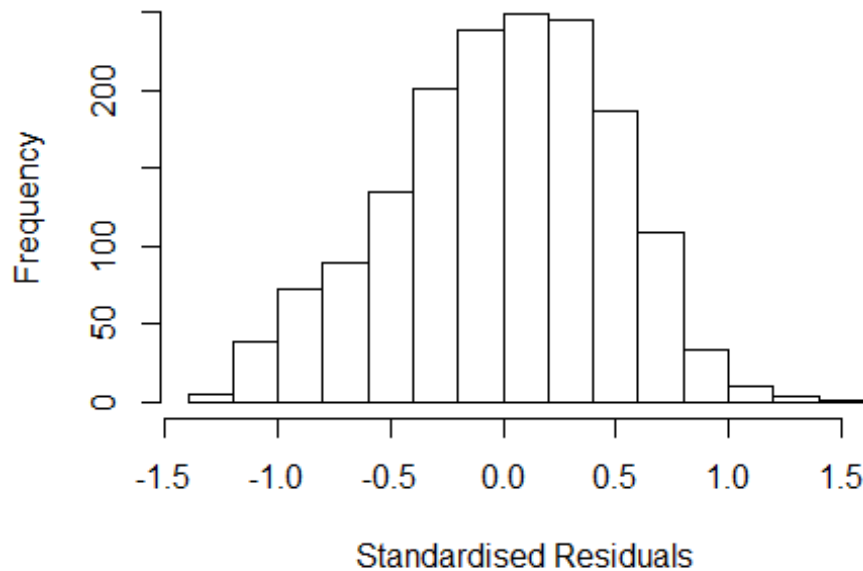


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.18, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 4.58128966415007"
## [1] "Male first author team size 2018 geometric mean: 3.74345560492839"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 860, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.13441270574067"
## [1] "Male last author team size 2018 geometric mean: 3.97311252307719"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 660, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.117 1      1.057
## LastAuthorFemale  1.098 1      1.048
## UniqueAuthors     1.219 4      1.025
## Year              1.296 16     1.008
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2906 -0.3430  0.0218  0.3396  1.4355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.0317    0.0938   11.00 < 2e-16 ***
## FirstAuthorFemale1  0.0235    0.0283    0.83  0.4068
## LastAuthorFemale1  0.0171    0.0322    0.53  0.5951
## UniqueAuthors2     0.0365    0.0511    0.72  0.4744
## UniqueAuthors3     0.1388    0.0487    2.85  0.0044 **
## UniqueAuthors4     0.1857    0.0452    4.10 4.3e-05 ***
## UniqueAuthors5     0.3393    0.0380    8.93 < 2e-16 ***
## Year1997           0.0840    0.1144    0.73  0.4630
## Year1998          -0.1402    0.1071   -1.31  0.1906
## Year1999          -0.0299    0.0983   -0.30  0.7608
```

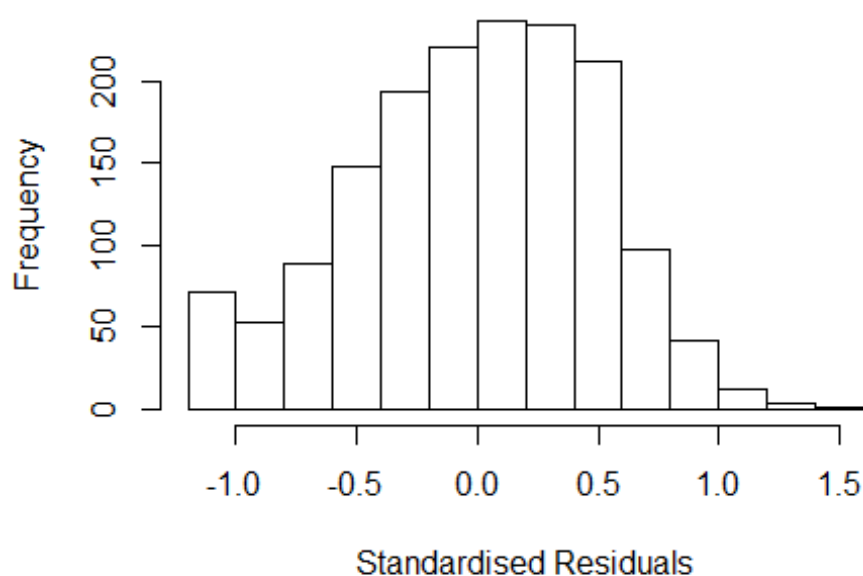
```

## Year2000          -0.1266      0.1041    -1.22    0.2240
## Year2001          -0.1275      0.1027    -1.24    0.2146
## Year2002          -0.0871      0.1054    -0.83    0.4087
## Year2003          -0.2363      0.1036    -2.28    0.0227 *
## Year2004          -0.1560      0.1023    -1.52    0.1275
## Year2005          -0.1987      0.1052    -1.89    0.0592 .
## Year2006          -0.1038      0.1082    -0.96    0.3374
## Year2007          -0.2432      0.1049    -2.32    0.0206 *
## Year2008          -0.2046      0.1062    -1.93    0.0542 .
## Year2009          -0.2134      0.1049    -2.04    0.0420 *
## Year2010          -0.1650      0.1051    -1.57    0.1166
## Year2011          -0.1822      0.1007    -1.81    0.0706 .
## Year2012          -0.3178      0.1025    -3.10    0.0020 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.497
## Multiple R-squared:  0.0838, Adjusted R-squared:  0.0712
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1497 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.384  0.875   0.949   0.912   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.18e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1      1.041
## LastAuthorFemale  1.069 1      1.034
## Year              1.082 16      1.002

```



## Residuals from first and last author



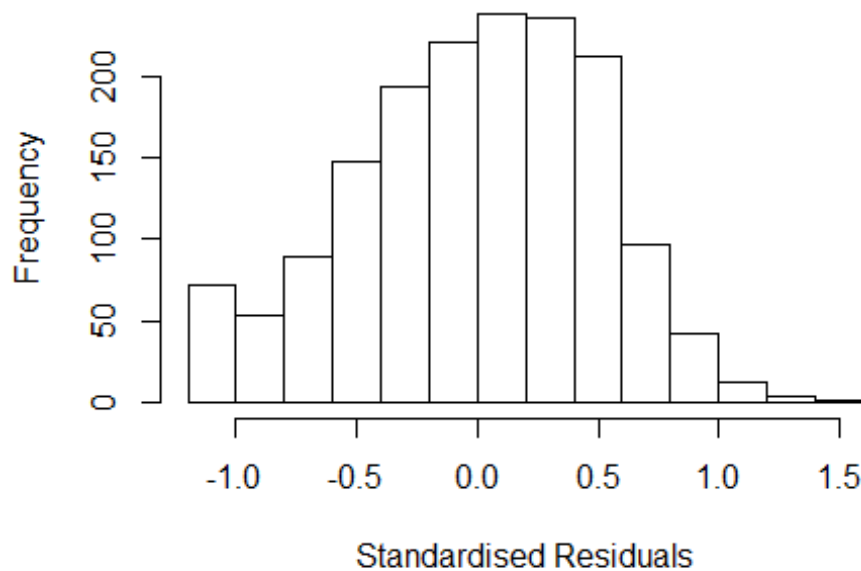
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1893 -0.3400 0.0246 0.3685 1.4117
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14812 0.09024 12.72 <2e-16 ***
## FirstAuthorFemale1 0.04116 0.02922 1.41 0.159
## LastAuthorFemale1 0.00344 0.03378 0.10 0.919
## Year1997 0.10921 0.11490 0.95 0.342
## Year1998 -0.06616 0.10797 -0.61 0.540
## Year1999 0.03125 0.09901 0.32 0.752
## Year2000 -0.02049 0.10629 -0.19 0.847
## Year2001 -0.06493 0.10339 -0.63 0.530
## Year2002 -0.01628 0.10510 -0.15 0.877
## Year2003 -0.12692 0.10233 -1.24 0.215
## Year2004 -0.06106 0.10141 -0.60 0.547
## Year2005 -0.12241 0.10746 -1.14 0.255
```

```

## Year2006      -0.04274    0.10985   -0.39    0.697
## Year2007      -0.15586    0.10499   -1.48    0.138
## Year2008      -0.11090    0.10688   -1.04    0.300
## Year2009      -0.13794    0.10632   -1.30    0.195
## Year2010      -0.09143    0.10671   -0.86    0.392
## Year2011      -0.11438    0.10033   -1.14    0.254
## Year2012      -0.22001    0.10212   -2.15    0.031 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0199, Adjusted R-squared:  0.00888
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1492 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.440  0.883   0.949   0.913   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.18e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1      1.023
## Year              1.047 16      1.001

```

## Residuals from first author



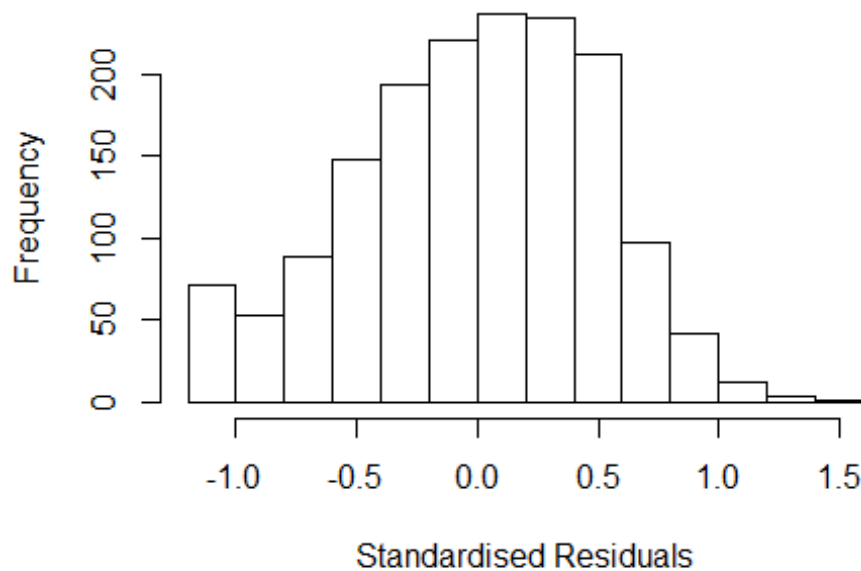
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1904 -0.3403 0.0236 0.3690 1.4106
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1486 0.0902 12.73 <2e-16 ***
## FirstAuthorFemale1 0.0417 0.0287 1.45 0.147
## Year1997 0.1093 0.1148 0.95 0.341
## Year1998 -0.0663 0.1080 -0.61 0.539
## Year1999 0.0310 0.0990 0.31 0.755
## Year2000 -0.0206 0.1063 -0.19 0.846
## Year2001 -0.0650 0.1034 -0.63 0.529
## Year2002 -0.0163 0.1051 -0.15 0.877
## Year2003 -0.1271 0.1023 -1.24 0.214
## Year2004 -0.0611 0.1014 -0.60 0.547
## Year2005 -0.1225 0.1075 -1.14 0.255
## Year2006 -0.0429 0.1099 -0.39 0.696
```

```

## Year2007          -0.1556      0.1049   -1.48    0.138
## Year2008          -0.1109      0.1069   -1.04    0.300
## Year2009          -0.1379      0.1063   -1.30    0.195
## Year2010          -0.0916      0.1067   -0.86    0.391
## Year2011          -0.1146      0.1004   -1.14    0.254
## Year2012          -0.2202      0.1021   -2.16    0.031 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0199, Adjusted R-squared:  0.0095
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1495 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.440  0.883  0.949  0.913  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.18e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1          1.017
## Year            1.033 16          1.001

```

## Residuals from last author



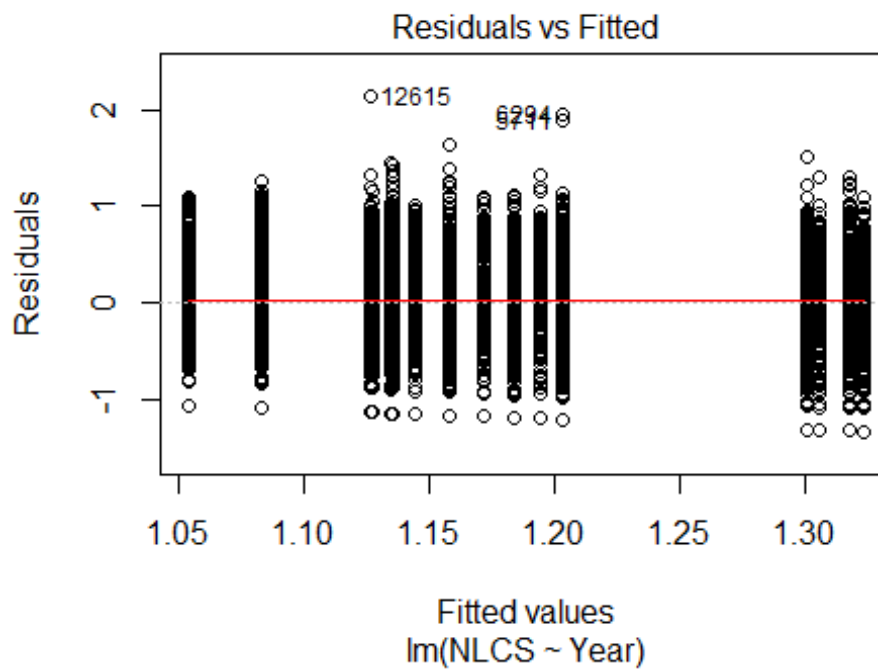
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1664 -0.3485 0.0275 0.3755 1.4386
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1530 0.0900 12.82 <2e-16 ***
## LastAuthorFemale1 0.0134 0.0330 0.41 0.685
## Year1997 0.1117 0.1142 0.98 0.328
## Year1998 -0.0612 0.1079 -0.57 0.571
## Year1999 0.0375 0.0983 0.38 0.703
## Year2000 -0.0150 0.1058 -0.14 0.887
## Year2001 -0.0609 0.1029 -0.59 0.554
## Year2002 -0.0120 0.1045 -0.11 0.909
## Year2003 -0.1188 0.1016 -1.17 0.242
## Year2004 -0.0585 0.1010 -0.58 0.562
## Year2005 -0.1128 0.1068 -1.06 0.291
## Year2006 -0.0338 0.1090 -0.31 0.757
```

```

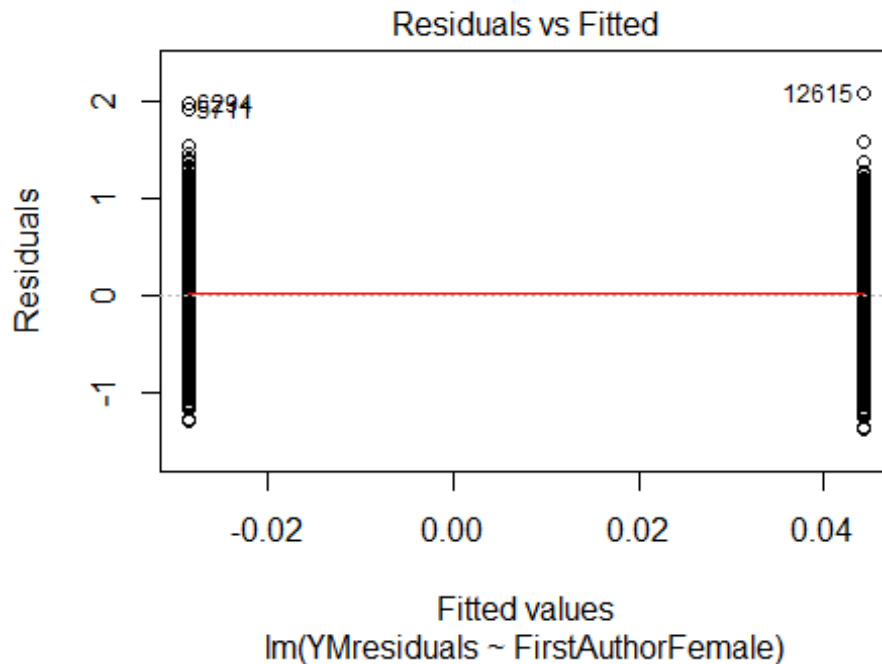
## Year2007          -0.1535      0.1047    -1.47      0.143
## Year2008          -0.1033      0.1065    -0.97      0.332
## Year2009          -0.1285      0.1053    -1.22      0.222
## Year2010          -0.0838      0.1063    -0.79      0.431
## Year2011          -0.1050      0.0994    -1.06      0.291
## Year2012          -0.2099      0.1012    -2.07      0.038 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.0186, Adjusted R-squared:  0.00815
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1486 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.423  0.883  0.949  0.913  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.18e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1617"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2728"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  774  815  700  737  802  886  769  699  701  797  861  798  810  850  861
## 2011 2012
##  886  865
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  271  269  279  346  314  313  372  360  385  472  495  489  455  541  592
## 2011 2012

```

```
## 615 630
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 235 231 243 287 262 264 307 291 308 372 414 418 382 468 520
## 2011 2012
## 544 555
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 55, df = 16, p-value = 3e-06
```



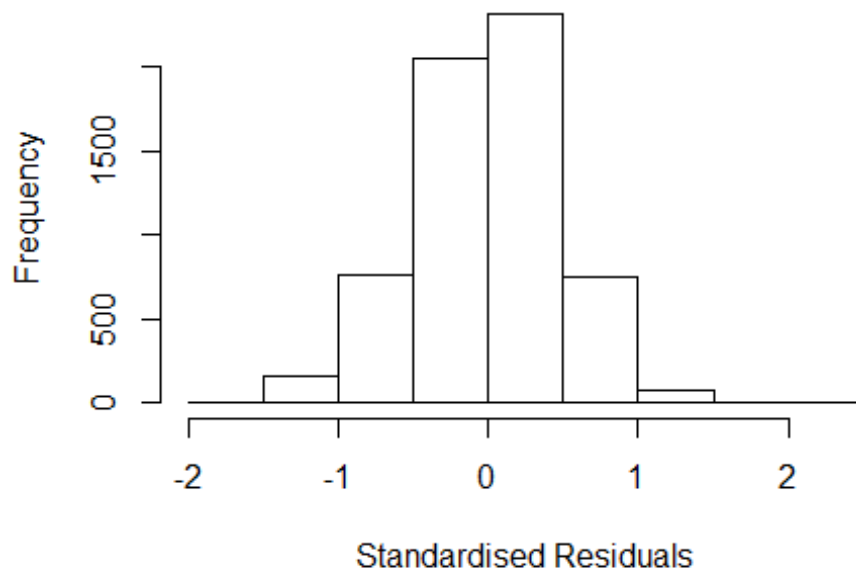
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 21, df = 1, p-value = 5e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.88417816953698"
## [1] "Male first author team size 2018 geometric mean: 4.43240852630525"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.59530681661772"
## [1] "Male last author team size 2018 geometric mean: 4.63657388246099"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1          1.014
## LastAuthorFemale  1.024 1          1.012
## UniqueAuthors    1.105 4          1.013
## Year              1.105 16         1.003
```



## Residuals from first and last author and team size



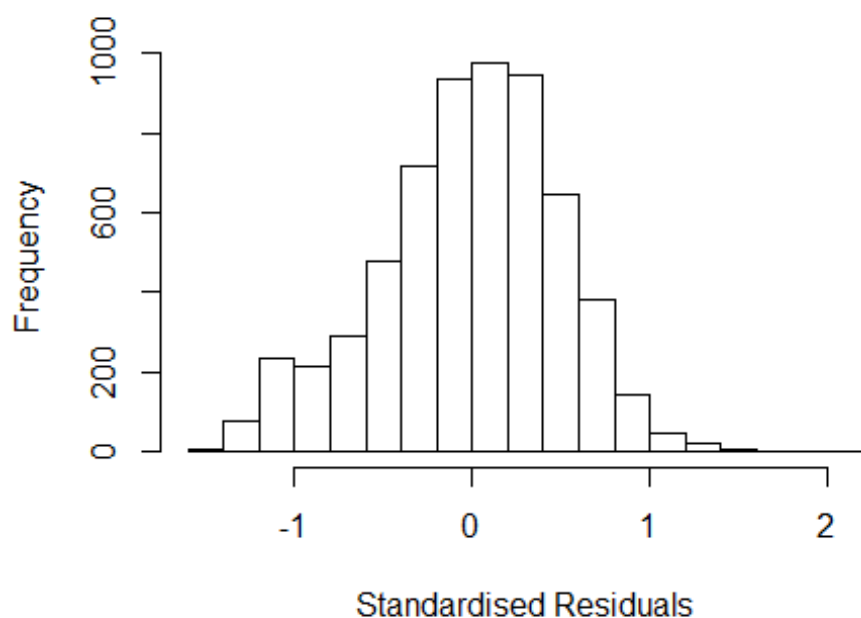
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5698 -0.3160 0.0205 0.3128 2.3514
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9324 0.0397 23.46 < 2e-16 ***
## FirstAuthorFemale1 0.0426 0.0124 3.42 0.00063 ***
## LastAuthorFemale1 0.0408 0.0140 2.91 0.00361 **
## UniqueAuthors2 0.3214 0.0281 11.44 < 2e-16 ***
## UniqueAuthors3 0.4203 0.0264 15.92 < 2e-16 ***
## UniqueAuthors4 0.4646 0.0264 17.58 < 2e-16 ***
## UniqueAuthors5 0.6203 0.0242 25.58 < 2e-16 ***
## Year1997 -0.1254 0.0474 -2.64 0.00822 **
## Year1998 -0.0255 0.0453 -0.56 0.57322
## Year1999 -0.0766 0.0453 -1.69 0.09086 .
```

```

## Year2000          -0.0692      0.0459   -1.51   0.13180
## Year2001          -0.1853      0.0450   -4.12   3.8e-05 ***
## Year2002          -0.1917      0.0413   -4.64   3.5e-06 ***
## Year2003          -0.1915      0.0414   -4.62   3.9e-06 ***
## Year2004          -0.2441      0.0422   -5.78   7.6e-09 ***
## Year2005          -0.3355      0.0429   -7.82   6.4e-15 ***
## Year2006          -0.2387      0.0405   -5.89   4.0e-09 ***
## Year2007          -0.2216      0.0404   -5.49   4.2e-08 ***
## Year2008          -0.2257      0.0410   -5.51   3.7e-08 ***
## Year2009          -0.2757      0.0407   -6.77   1.4e-11 ***
## Year2010          -0.2413      0.0393   -6.13   9.3e-10 ***
## Year2011          -0.3110      0.0397   -7.84   5.2e-15 ***
## Year2012          -0.2850      0.0411   -6.94   4.5e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.462
## Multiple R-squared:  0.174, Adjusted R-squared:  0.171
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 2 observations c(1580,4084) are outliers with |weight| = 0 ( < 1.6e-05);
## 445 weights are ~= 1. The remaining 5654 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0966 0.8680 0.9510 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.009 1          1.005
## LastAuthorFemale 1.008 1          1.004
## Year          1.017 16          1.001

```

## Residuals from first and last author

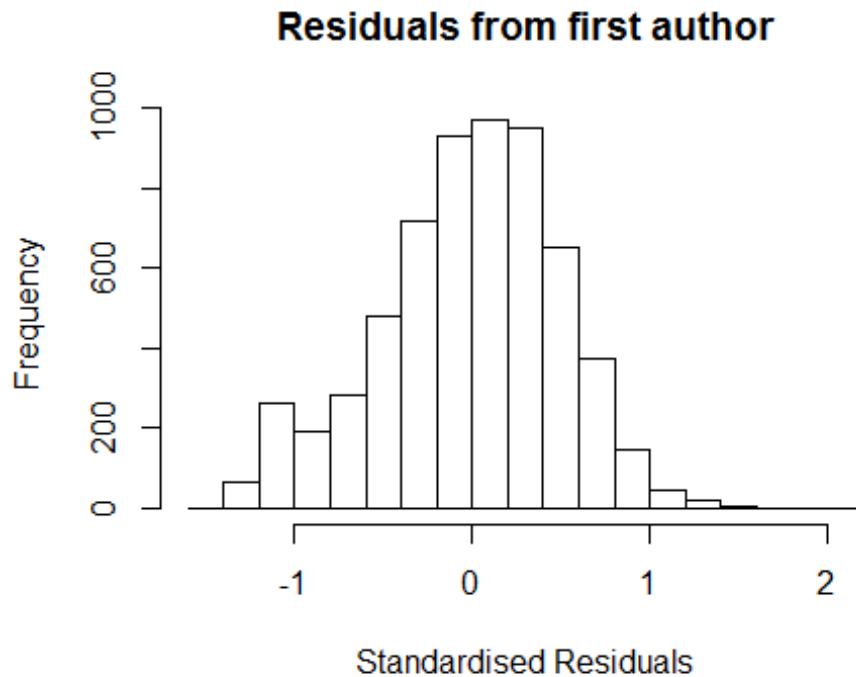


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4043 -0.3270  0.0193  0.3323  2.0828
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3274     0.0341   38.96 < 2e-16 ***
## FirstAuthorFemale1  0.0753     0.0132    5.70 1.2e-08 ***
## LastAuthorFemale1  0.0200     0.0152    1.32 0.18712
## Year1997          -0.1187     0.0517   -2.30 0.02164 *
## Year1998          -0.0184     0.0453   -0.41 0.68471
## Year1999          -0.0537     0.0456   -1.18 0.23875
## Year2000          -0.0458     0.0466   -0.98 0.32575
## Year2001          -0.1742     0.0468   -3.72 0.00020 ***
## Year2002          -0.1620     0.0429   -3.78 0.00016 ***
## Year2003          -0.1538     0.0425   -3.62 0.00030 ***
## Year2004          -0.2206     0.0435   -5.06 4.2e-07 ***
## Year2005          -0.3294     0.0464   -7.10 1.4e-12 ***
```

```

## Year2006          -0.2207      0.0422    -5.23  1.8e-07 ***
## Year2007          -0.1699      0.0408    -4.17  3.1e-05 ***
## Year2008          -0.2099      0.0417    -5.03  5.0e-07 ***
## Year2009          -0.2255      0.0411    -5.48  4.4e-08 ***
## Year2010          -0.1747      0.0401    -4.36  1.3e-05 ***
## Year2011          -0.2654      0.0407    -6.53  7.2e-11 ***
## Year2012          -0.2039      0.0421    -4.84  1.3e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.482
## Multiple R-squared:  0.0297, Adjusted R-squared:  0.0268
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 509 weights are ~= 1. The remaining 5592 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0228 0.8640 0.9490 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.009 1      1.004
## Year      1.009 16      1.000

```

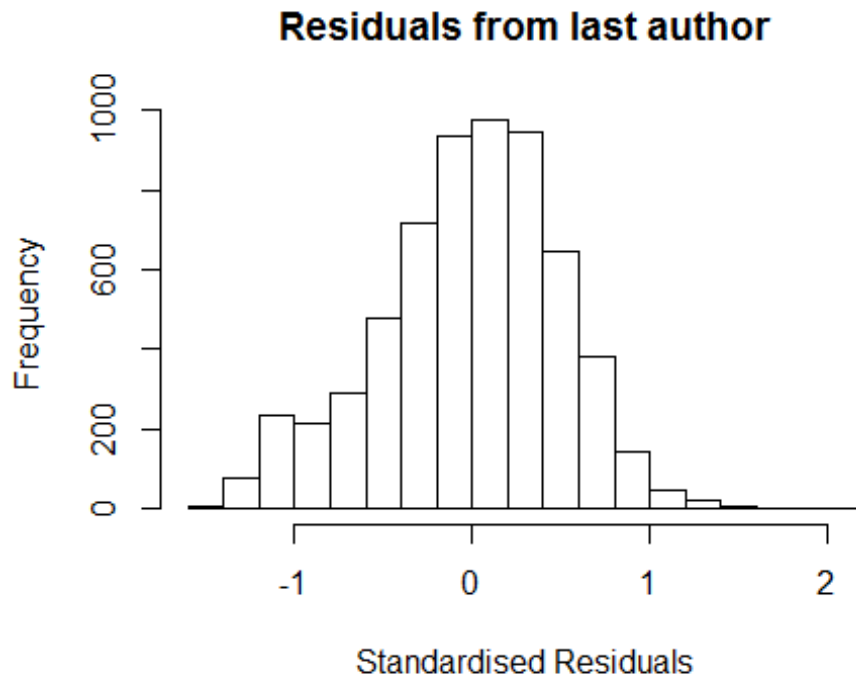


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.408 -0.329  0.022  0.333  2.077
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3302     0.0339   39.26 < 2e-16 ***
## FirstAuthorFemale1  0.0776     0.0133    5.84 5.4e-09 ***
## Year1997         -0.1184     0.0516   -2.29 0.02178 *
## Year1998         -0.0176     0.0453   -0.39 0.69767
## Year1999         -0.0537     0.0456   -1.18 0.23847
## Year2000         -0.0455     0.0465   -0.98 0.32787
## Year2001         -0.1729     0.0468   -3.69 0.00022 ***
## Year2002         -0.1606     0.0428   -3.75 0.00018 ***
## Year2003         -0.1528     0.0425   -3.60 0.00032 ***
## Year2004         -0.2190     0.0435   -5.04 4.8e-07 ***
## Year2005         -0.3272     0.0463   -7.07 1.8e-12 ***
## Year2006         -0.2185     0.0422   -5.18 2.3e-07 ***
```

```

## Year2007          -0.1688      0.0407   -4.15   3.4e-05 ***
## Year2008          -0.2094      0.0417   -5.02   5.2e-07 ***
## Year2009          -0.2246      0.0411   -5.47   4.8e-08 ***
## Year2010          -0.1740      0.0401   -4.34   1.4e-05 ***
## Year2011          -0.2647      0.0406   -6.52   7.8e-11 ***
## Year2012          -0.2032      0.0420   -4.83   1.4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.0294, Adjusted R-squared:  0.0266
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 498 weights are ~= 1. The remaining 5603 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0245 0.8650 0.9490 0.8990 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.008 1          1.004
## Year              1.008 16          1.000

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3655 -0.3297 0.0261 0.3359 2.1297
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3497 0.0337 40.02 < 2e-16 ***
## LastAuthorFemale1 0.0318 0.0153 2.08 0.03747 *
## Year1997 -0.1165 0.0514 -2.27 0.02345 *
## Year1998 -0.0159 0.0456 -0.35 0.72677
## Year1999 -0.0457 0.0455 -1.00 0.31565
## Year2000 -0.0440 0.0464 -0.95 0.34344
## Year2001 -0.1715 0.0467 -3.67 0.00024 ***
## Year2002 -0.1578 0.0430 -3.67 0.00024 ***
## Year2003 -0.1498 0.0427 -3.51 0.00045 ***
## Year2004 -0.2137 0.0435 -4.91 9.3e-07 ***
## Year2005 -0.3226 0.0465 -6.94 4.2e-12 ***
## Year2006 -0.2156 0.0422 -5.11 3.3e-07 ***
```

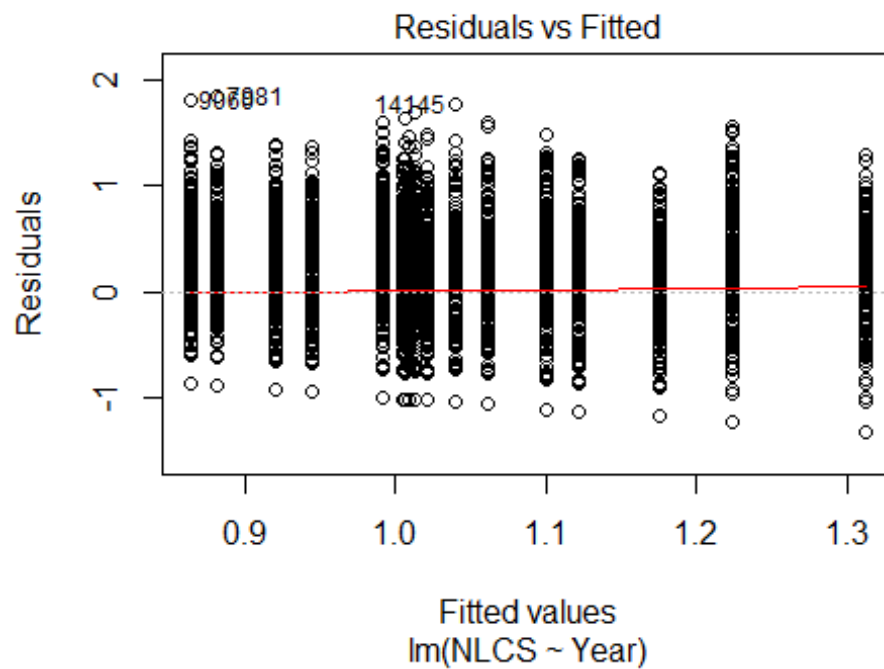
```

## Year2007          -0.1644      0.0407    -4.04  5.5e-05 ***
## Year2008          -0.2030      0.0418    -4.86  1.2e-06 ***
## Year2009          -0.2194      0.0412    -5.32  1.1e-07 ***
## Year2010          -0.1675      0.0401    -4.17  3.0e-05 ***
## Year2011          -0.2593      0.0407    -6.36  2.1e-10 ***
## Year2012          -0.1932      0.0420    -4.60  4.4e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.0245, Adjusted R-squared:  0.0218
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 489 weights are ~ = 1. The remaining 5612 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0136 0.8640 0.9490 0.8990 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.64e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6101"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2729"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 779 792 713 815 773 798 735 769 704 777 887 865 784 765 710
## 2011 2012
## 637 549
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 270 280 231 335 284 222 302 299 301 372 368 334 328 321 343
## 2011 2012

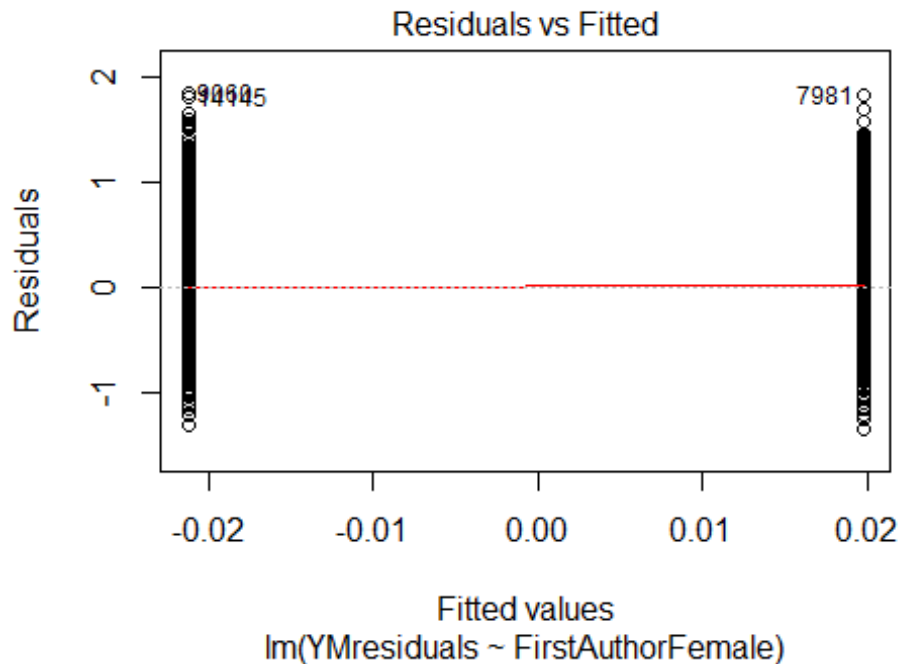
```



```
## 351 293
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 238 249 197 295 257 205 241 236 219 313 316 290 273 277 297
## 2011 2012
## 304 263
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 66, df = 16, p-value = 5e-08
```

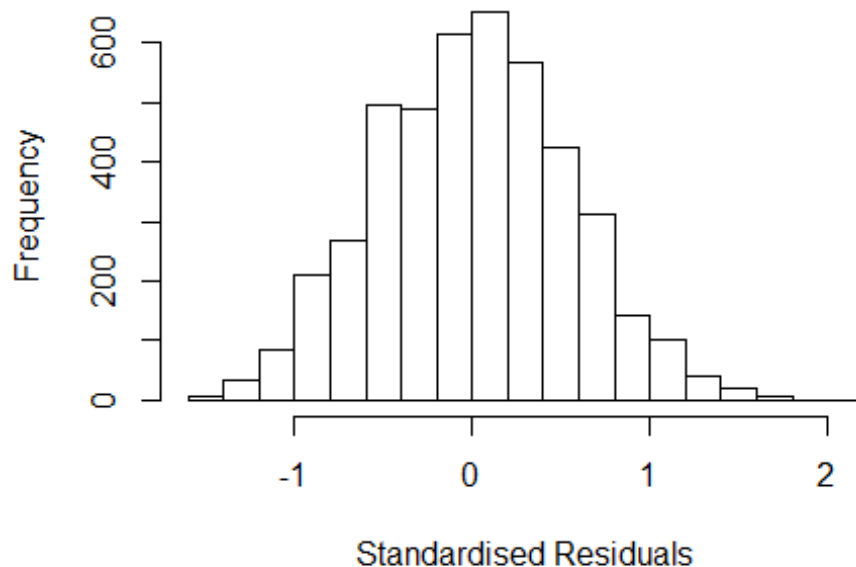


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 3.43896956114093"
## [1] "Male first author team size 2018 geometric mean: 3.63211448415753"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.21314511244817"
## [1] "Male last author team size 2018 geometric mean: 3.86711115495986"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6900, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.136 1          1.066
## LastAuthorFemale  1.138 1          1.067
## UniqueAuthors     1.155 4          1.018
## Year              1.200 16          1.006
```

## Residuals from first and last author and team size



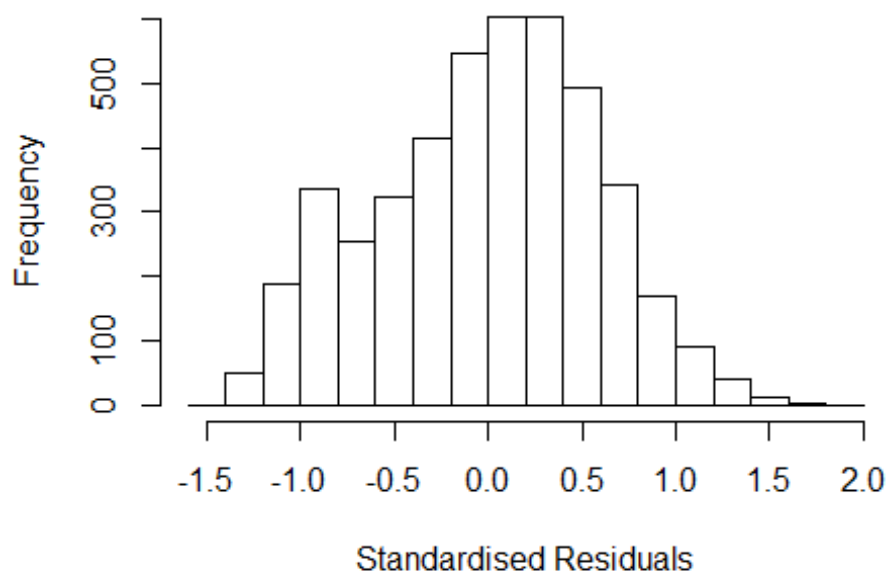
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.58450 -0.38915 0.00847 0.37367 2.07078
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7102 0.0488 14.56 < 2e-16 ***
## FirstAuthorFemale1 0.0600 0.0180 3.34 0.00086 ***
## LastAuthorFemale1 0.0283 0.0188 1.51 0.13162
## UniqueAuthors2 0.3267 0.0322 10.15 < 2e-16 ***
## UniqueAuthors3 0.4665 0.0310 15.06 < 2e-16 ***
## UniqueAuthors4 0.6054 0.0302 20.08 < 2e-16 ***
## UniqueAuthors5 0.6969 0.0275 25.37 < 2e-16 ***
## Year1997 0.0926 0.0612 1.51 0.13002
## Year1998 0.1774 0.0630 2.82 0.00489 **
## Year1999 -0.1633 0.0572 -2.85 0.00436 **
```

```

## Year2000          -0.0310      0.0587   -0.53  0.59714
## Year2001          -0.0648      0.0605   -1.07  0.28446
## Year2002          -0.2260      0.0560   -4.03  5.6e-05 ***
## Year2003          -0.1514      0.0545   -2.78  0.00544 **
## Year2004          -0.1873      0.0546   -3.43  0.00060 ***
## Year2005          -0.2681      0.0531   -5.05  4.5e-07 ***
## Year2006          -0.2888      0.0538   -5.37  8.4e-08 ***
## Year2007          -0.2715      0.0541   -5.02  5.5e-07 ***
## Year2008          -0.2397      0.0546   -4.39  1.2e-05 ***
## Year2009          -0.1899      0.0564   -3.37  0.00077 ***
## Year2010          -0.2055      0.0549   -3.74  0.00019 ***
## Year2011          -0.1570      0.0541   -2.90  0.00371 **
## Year2012          -0.1764      0.0569   -3.10  0.00193 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.207, Adjusted R-squared:  0.203
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 367 weights are ~= 1. The remaining 4103 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.105  0.871  0.944  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.24e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.100 1          1.049
## LastAuthorFemale  1.085 1          1.041
## Year              1.068 16          1.002

```

## Residuals from first and last author



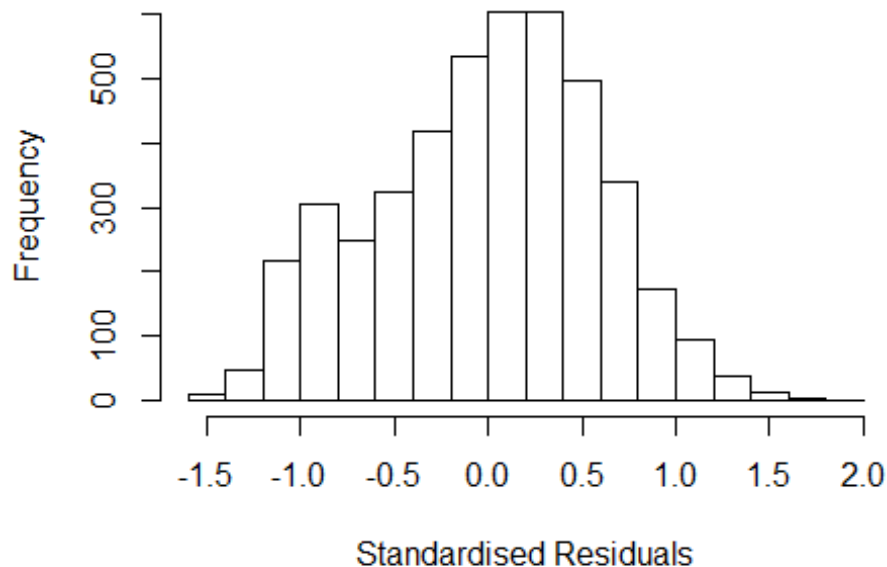
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4489 -0.4199 0.0434 0.4112 1.8414
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11465 0.04895 22.77 < 2e-16 ***
## FirstAuthorFemale1 0.08353 0.01948 4.29 1.8e-05 ***
## LastAuthorFemale1 -0.05928 0.02034 -2.91 0.00359 **
## Year1997 0.15639 0.06692 2.34 0.01948 *
## Year1998 0.25076 0.06561 3.82 0.00013 ***
## Year1999 -0.14961 0.06620 -2.26 0.02388 *
## Year2000 0.00743 0.06113 0.12 0.90325
## Year2001 0.04956 0.06292 0.79 0.43098
## Year2002 -0.14844 0.06089 -2.44 0.01481 *
## Year2003 -0.06116 0.05952 -1.03 0.30421
## Year2004 -0.09897 0.05999 -1.65 0.09907 .
## Year2005 -0.25535 0.06072 -4.21 2.7e-05 ***
```

```

## Year2006      -0.29642    0.06017   -4.93  8.7e-07 ***
## Year2007      -0.24757    0.06147   -4.03  5.7e-05 ***
## Year2008      -0.20630    0.06034   -3.42  0.00063 ***
## Year2009      -0.12477    0.06077   -2.05  0.04013 *
## Year2010      -0.13353    0.06000   -2.23  0.02610 *
## Year2011      -0.06121    0.05902   -1.04  0.29973
## Year2012      -0.08573    0.06092   -1.41  0.15946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0506, Adjusted R-squared:  0.0467
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 366 weights are ~= 1. The remaining 4104 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.325  0.864  0.948  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.045 1      1.022
## Year      1.045 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4225 -0.4226 0.0398 0.4161 1.8095
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10406 0.04921 22.44 < 2e-16 ***
## FirstAuthorFemale1 0.06551 0.01909 3.43 0.00061 ***
## Year1997 0.15497 0.06718 2.31 0.02111 *
## Year1998 0.25296 0.06589 3.84 0.00013 ***
## Year1999 -0.14822 0.06661 -2.22 0.02613 *
## Year2000 0.00933 0.06152 0.15 0.87953
## Year2001 0.05072 0.06330 0.80 0.42298
## Year2002 -0.14645 0.06127 -2.39 0.01687 *
## Year2003 -0.06138 0.05994 -1.02 0.30587
## Year2004 -0.10034 0.06050 -1.66 0.09725 .
## Year2005 -0.25704 0.06106 -4.21 2.6e-05 ***
## Year2006 -0.29962 0.06045 -4.96 7.4e-07 ***
```

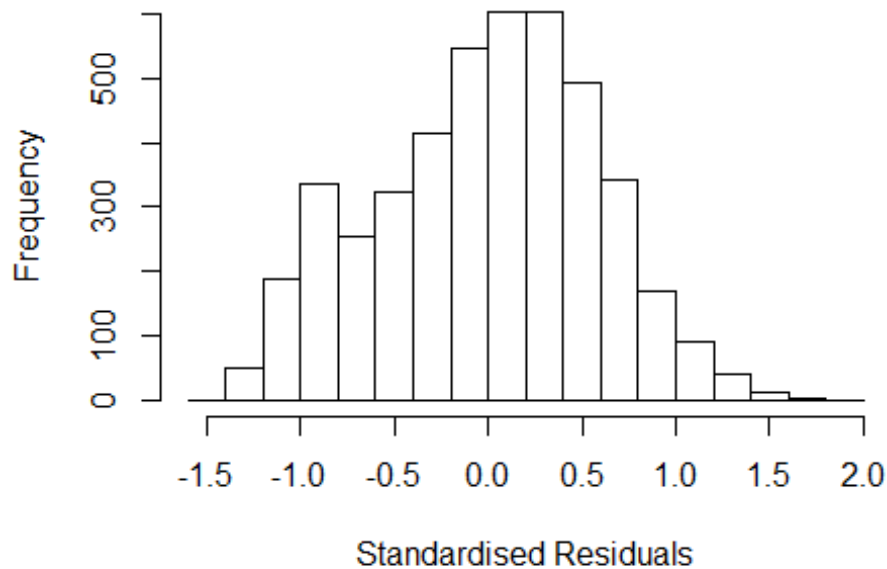
```

## Year2007          -0.24643    0.06185   -3.98  6.9e-05 ***
## Year2008          -0.21078    0.06064   -3.48  0.00051 ***
## Year2009          -0.12418    0.06118   -2.03  0.04244 *
## Year2010          -0.13762    0.06024   -2.28  0.02239 *
## Year2011          -0.06442    0.05935   -1.09  0.27779
## Year2012          -0.09119    0.06105   -1.49  0.13533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.0489, Adjusted R-squared:  0.0453
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 389 weights are ~= 1. The remaining 4081 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.341  0.865  0.947  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1          1.014
## Year            1.028 16          1.001

```



## Residuals from last author



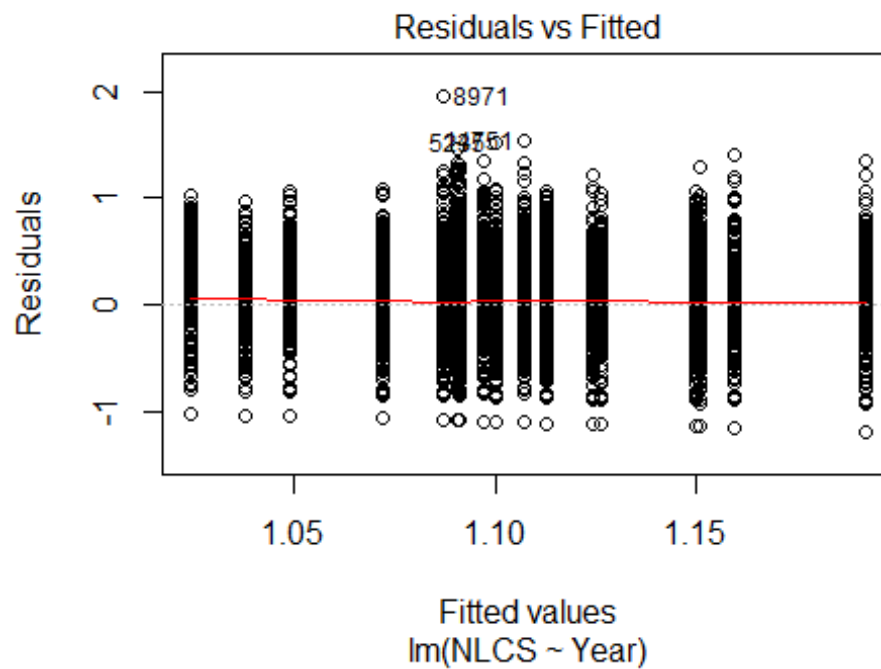
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3928 -0.4192 0.0429 0.4194 1.8289
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1390 0.0482 23.62 < 2e-16 ***
## LastAuthorFemale1 -0.0319 0.0199 -1.61 0.10833
## Year1997 0.1508 0.0664 2.27 0.02324 *
## Year1998 0.2538 0.0653 3.89 0.00010 ***
## Year1999 -0.1440 0.0661 -2.18 0.02942 *
## Year2000 0.0119 0.0608 0.20 0.84480
## Year2001 0.0544 0.0627 0.87 0.38607
## Year2002 -0.1377 0.0608 -2.26 0.02361 *
## Year2003 -0.0532 0.0594 -0.90 0.37039
## Year2004 -0.0865 0.0597 -1.45 0.14733
## Year2005 -0.2459 0.0602 -4.09 4.4e-05 ***
## Year2006 -0.2865 0.0598 -4.79 1.7e-06 ***
```

```

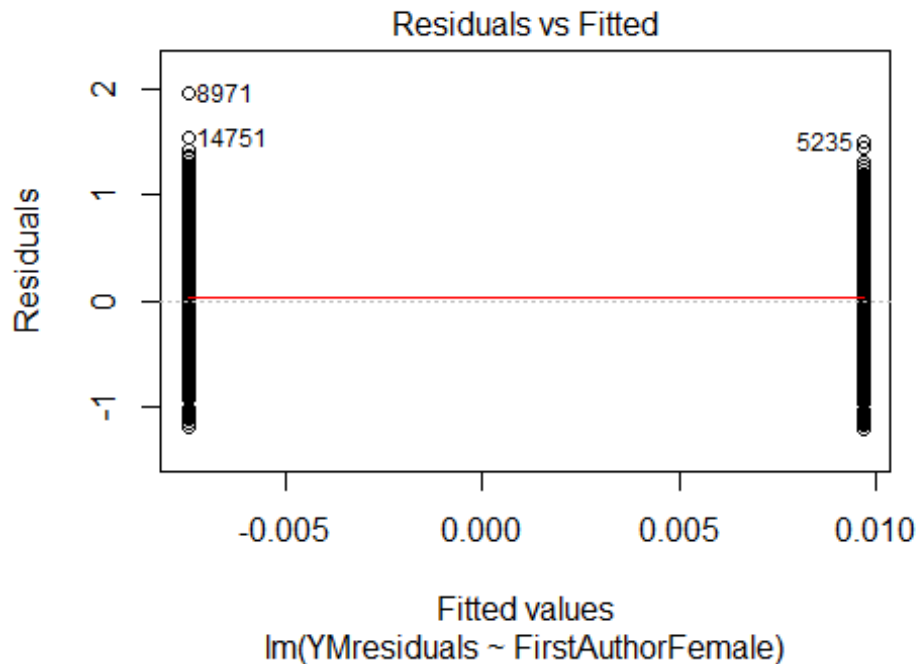
## Year2007          -0.2329      0.0610    -3.82   0.00014 ***
## Year2008          -0.1968      0.0601    -3.28   0.00106 **
## Year2009          -0.1073      0.0606    -1.77   0.07658 .
## Year2010          -0.1214      0.0595    -2.04   0.04139 *
## Year2011          -0.0461      0.0589    -0.78   0.43301
## Year2012          -0.0697      0.0606    -1.15   0.24987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.0465, Adjusted R-squared:  0.0428
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 351 weights are ~= 1. The remaining 4119 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.335  0.868  0.948  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.24e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4470"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2730"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 730 720 634 677 770 727 689 621 687 701 836 846 845 853 800
## 2011 2012
## 781 775
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 219 188 226 247 196 287 261 301 344 396 445 439 442 443
## 2011 2012

```

```
## 434 417
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 185 199 167 198 223 183 243 207 236 280 339 383 374 373 376
## 2011 2012
## 380 369
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 64, df = 16, p-value = 1e-07
```

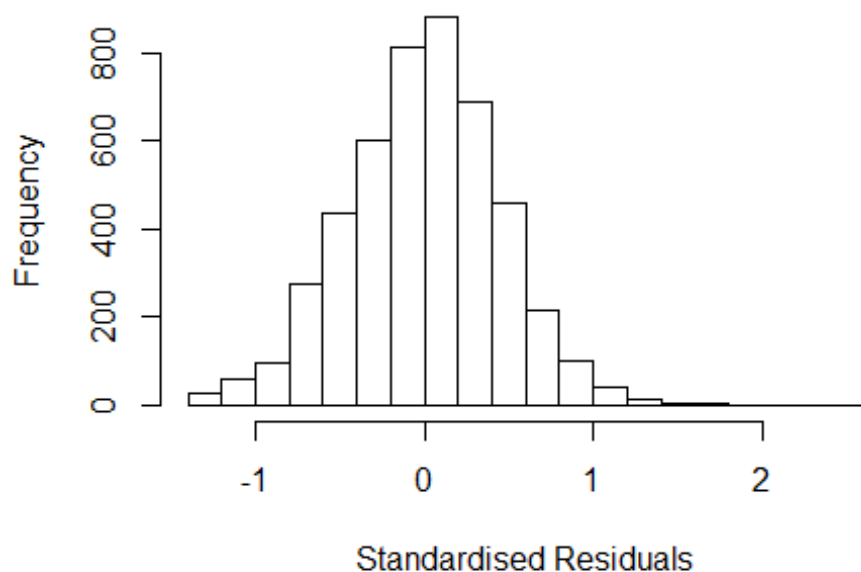


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 5e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 5.69462669300661"
## [1] "Male first author team size 2018 geometric mean: 5.96602850543858"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.9213431533631"
## [1] "Male last author team size 2018 geometric mean: 5.80018431906106"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.036 1          1.018
## LastAuthorFemale  1.040 1          1.020
## UniqueAuthors    1.100 4          1.012
## Year             1.115 16          1.003
```

## Residuals from first and last author and team size



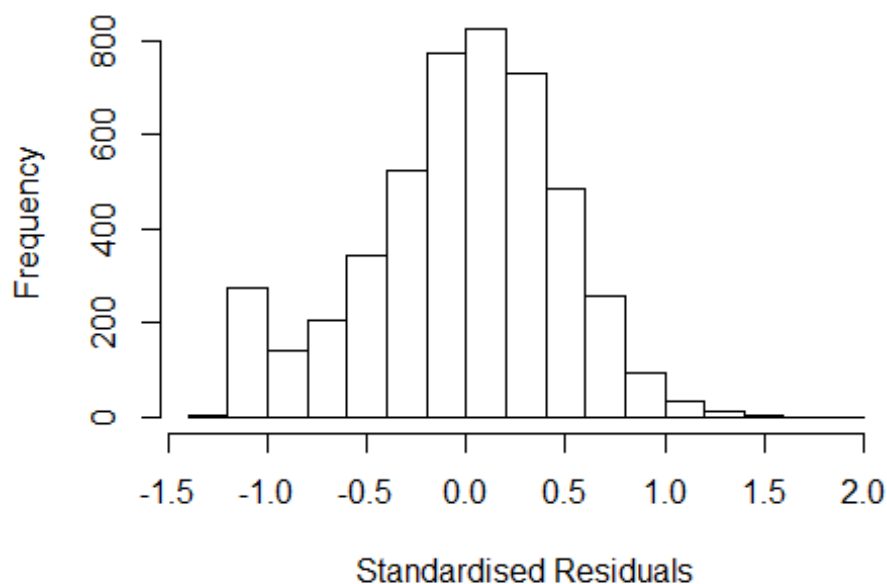
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.397 -0.292 0.016 0.295 2.439
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.74041 0.04406 16.81 < 2e-16 ***
## FirstAuthorFemale1 -0.00431 0.01347 -0.32 0.74931
## LastAuthorFemale1 0.01453 0.01506 0.96 0.33480
## UniqueAuthors2 0.36043 0.03593 10.03 < 2e-16 ***
## UniqueAuthors3 0.43613 0.03355 13.00 < 2e-16 ***
## UniqueAuthors4 0.47104 0.03364 14.00 < 2e-16 ***
## UniqueAuthors5 0.66069 0.02846 23.22 < 2e-16 ***
## Year1997 0.00578 0.05177 0.11 0.91111
## Year1998 -0.03244 0.05193 -0.62 0.53218
## Year1999 -0.09045 0.04788 -1.89 0.05895 .
```

```

## Year2000      -0.15775      0.04994      -3.16      0.00159 **
## Year2001      -0.11485      0.05479      -2.10      0.03611 *
## Year2002      -0.11117      0.04765      -2.33      0.01970 *
## Year2003      -0.15958      0.04638      -3.44      0.00058 ***
## Year2004      -0.17249      0.04576      -3.77      0.00017 ***
## Year2005      -0.11523      0.04499      -2.56      0.01047 *
## Year2006      -0.13814      0.04438      -3.11      0.00186 **
## Year2007      -0.15624      0.04444      -3.52      0.00044 ***
## Year2008      -0.11566      0.04449      -2.60      0.00935 **
## Year2009      -0.10549      0.04339      -2.43      0.01508 *
## Year2010      -0.15474      0.04467      -3.46      0.00054 ***
## Year2011      -0.14464      0.04506      -3.21      0.00134 **
## Year2012      -0.19197      0.04547      -4.22      2.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.43
## Multiple R-squared:  0.197, Adjusted R-squared:  0.193
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 2295 is an outlier with |weight| = 0 ( < 2.1e-05);
## 390 weights are ~= 1. The remaining 4324 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0923 0.8550 0.9490 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.12e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.019 1 1.009
## LastAuthorFemale 1.032 1 1.016
## Year 1.050 16 1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2037 -0.3080 0.0218 0.3191 1.9327
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16667 0.04120 28.32 <2e-16 ***
## FirstAuthorFemale1 0.02400 0.01451 1.65 0.098 .
## LastAuthorFemale1 -0.02856 0.01655 -1.73 0.084 .
## Year1997 0.03703 0.05618 0.66 0.510
## Year1998 0.00287 0.05457 0.05 0.958
## Year1999 -0.03349 0.05134 -0.65 0.514
## Year2000 -0.11234 0.06199 -1.81 0.070 .
## Year2001 -0.08976 0.06006 -1.49 0.135
## Year2002 -0.05631 0.05276 -1.07 0.286
## Year2003 -0.09156 0.05003 -1.83 0.067 .
## Year2004 -0.07590 0.04963 -1.53 0.126
## Year2005 -0.01272 0.04853 -0.26 0.793
```

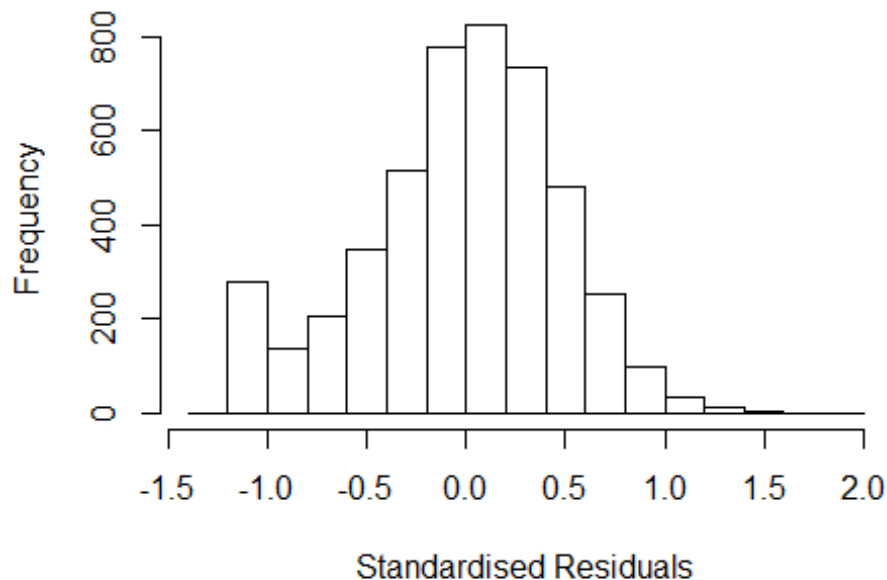
```

## Year2006      -0.05841    0.04747   -1.23    0.219
## Year2007      -0.05502    0.04876   -1.13    0.259
## Year2008      -0.02736    0.04808   -0.57    0.569
## Year2009       0.00447    0.04626    0.10    0.923
## Year2010      -0.03643    0.04825   -0.76    0.450
## Year2011      -0.03149    0.04794   -0.66    0.511
## Year2012      -0.05901    0.04946   -1.19    0.233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.00612,    Adjusted R-squared:  0.00231
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 379 weights are ~= 1. The remaining 4336 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0369 0.8600 0.9490 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.12e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.021 1      1.010
## Year      1.021 16      1.001

```



## Residuals from first author



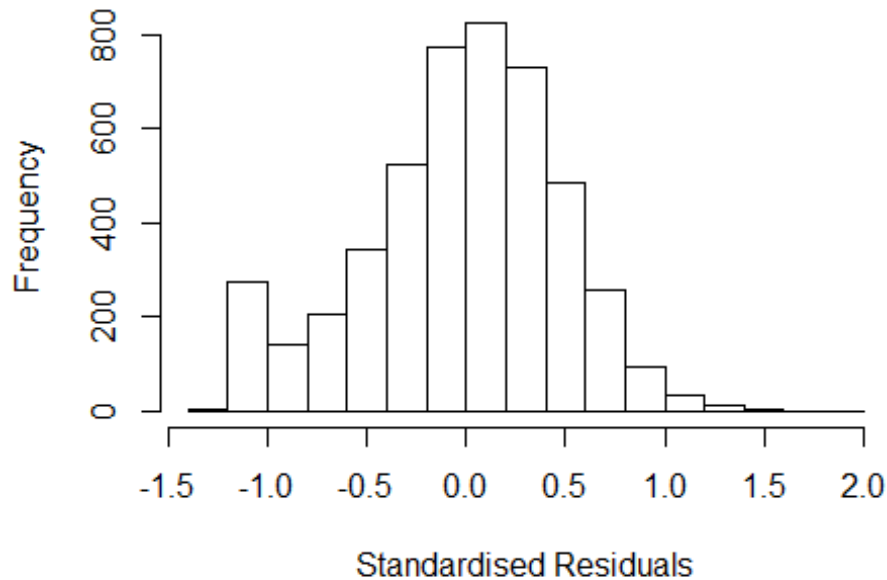
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2190 -0.3057 0.0215 0.3177 1.9373
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16121 0.04108 28.27 <2e-16 ***
## FirstAuthorFemale1 0.01964 0.01468 1.34 0.181
## Year1997 0.03814 0.05630 0.68 0.498
## Year1998 0.00481 0.05446 0.09 0.930
## Year1999 -0.03281 0.05135 -0.64 0.523
## Year2000 -0.11114 0.06220 -1.79 0.074 .
## Year2001 -0.08913 0.06015 -1.48 0.138
## Year2002 -0.05457 0.05280 -1.03 0.301
## Year2003 -0.09173 0.05013 -1.83 0.067 .
## Year2004 -0.07663 0.04978 -1.54 0.124
## Year2005 -0.01366 0.04854 -0.28 0.778
## Year2006 -0.05749 0.04753 -1.21 0.227
```

```

## Year2007          -0.05571    0.04881   -1.14    0.254
## Year2008          -0.02697    0.04820   -0.56    0.576
## Year2009           0.00427    0.04632    0.09    0.926
## Year2010          -0.03801    0.04828   -0.79    0.431
## Year2011          -0.03223    0.04801   -0.67    0.502
## Year2012          -0.06010    0.04948   -1.21    0.225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.00546,    Adjusted R-squared:  0.00186
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 403 weights are ~= 1. The remaining 4312 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0353 0.8580 0.9480 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.12e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.031 1          1.016
## Year              1.031 16          1.001

```

## Residuals from last author



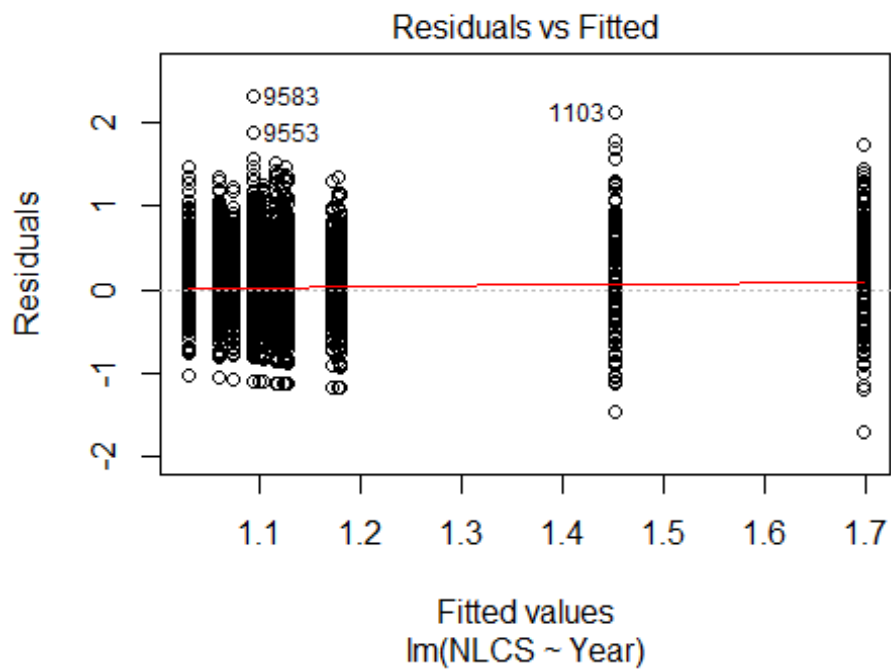
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2124 -0.3046  0.0236  0.3176  1.9227
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.18e+00  4.05e-02  29.06  <2e-16 ***
## LastAuthorFemale1 -2.39e-02  1.67e-02  -1.43   0.154
## Year1997         3.61e-02  5.60e-02   0.64   0.520
## Year1998          9.62e-05  5.45e-02   0.00   0.999
## Year1999        -3.34e-02  5.11e-02  -0.65   0.514
## Year2000        -1.13e-01  6.18e-02  -1.83   0.067 .
## Year2001        -8.97e-02  5.99e-02  -1.50   0.134
## Year2002        -5.59e-02  5.26e-02  -1.06   0.289
## Year2003        -9.16e-02  4.99e-02  -1.83   0.067 .
## Year2004        -7.65e-02  4.95e-02  -1.54   0.123
## Year2005        -1.28e-02  4.84e-02  -0.26   0.792
## Year2006        -5.81e-02  4.73e-02  -1.23   0.220
```

```

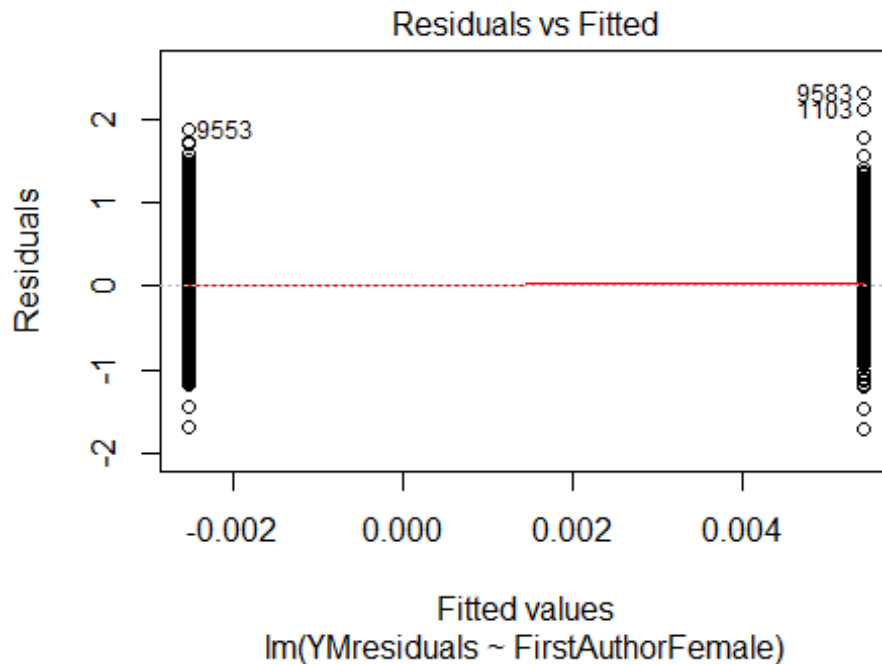
## Year2007          -5.49e-02   4.86e-02   -1.13   0.259
## Year2008          -2.71e-02   4.80e-02   -0.56   0.573
## Year2009           5.45e-03   4.61e-02    0.12   0.906
## Year2010          -3.54e-02   4.81e-02   -0.74   0.462
## Year2011          -3.05e-02   4.78e-02   -0.64   0.524
## Year2012          -5.92e-02   4.93e-02   -1.20   0.230
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.00561,    Adjusted R-squared:  0.00202
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 374 weights are ~= 1. The remaining 4341 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0401 0.8600 0.9490 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.12e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4715"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2731"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 623 568 390 390 448 449 456 373 408 453 505 567 491 524 520
## 2011 2012
## 512 485
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 178 179 196 227 199 163 267 221 212 253 309 352 285 320 349
## 2011 2012

```

```
## 335 327
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 161 158 178 195 179 137 209 174 178 197 237 295 244 268 291
## 2011 2012
## 277 262
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 220, df = 16, p-value <2e-16
```

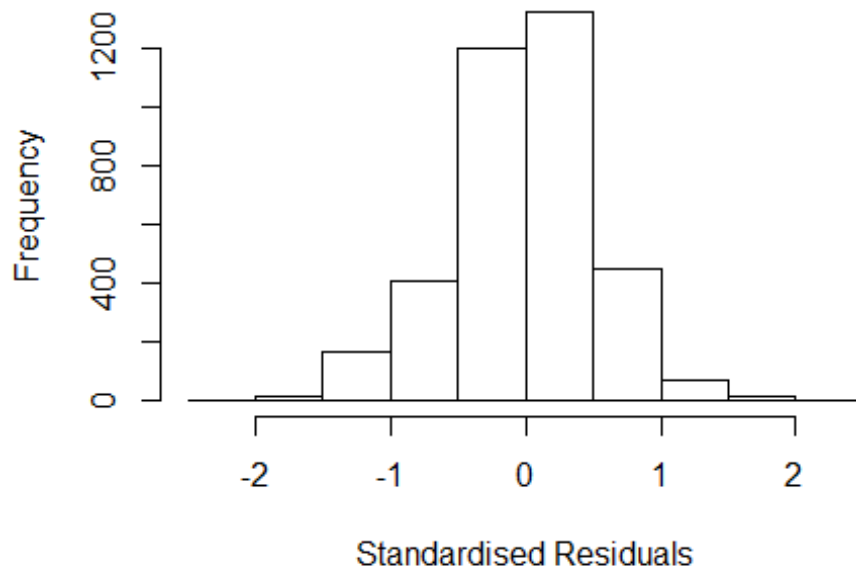


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.0623660216468"
## [1] "Male first author team size 2018 geometric mean: 3.68028012963914"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.12142332325185"
## [1] "Male last author team size 2018 geometric mean: 3.74582561568854"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1 1.019
## LastAuthorFemale 1.050 1 1.024
## UniqueAuthors 1.117 4 1.014
## Year 1.150 16 1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -2.0221 -0.3205 0.0168 0.3277 2.1387
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.51072 0.08669 17.43 < 2e-16 ***
## FirstAuthorFemale1 -0.00578 0.01802 -0.32 0.749
## LastAuthorFemale1 -0.02661 0.02136 -1.25 0.213
## UniqueAuthors2 0.25997 0.03849 6.75 1.7e-11 ***
## UniqueAuthors3 0.32854 0.03784 8.68 < 2e-16 ***
## UniqueAuthors4 0.36476 0.03973 9.18 < 2e-16 ***
## UniqueAuthors5 0.51142 0.03768 13.57 < 2e-16 ***
## Year1997 -0.25484 0.09977 -2.55 0.011 *
## Year1998 -0.66616 0.08728 -7.63 2.9e-14 ***
## Year1999 -0.57592 0.08434 -6.83 1.0e-11 ***
```

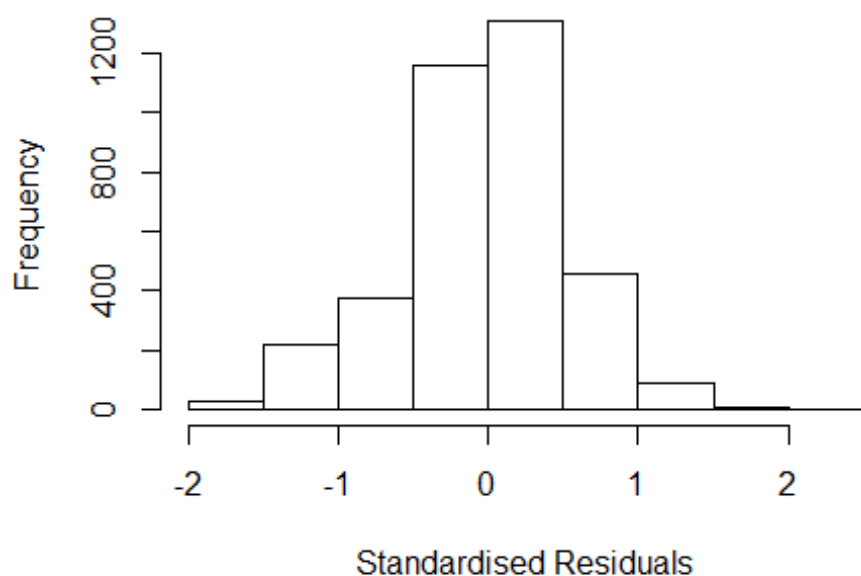
```

## Year2000      -0.59290    0.08793    -6.74    1.8e-11 ***
## Year2001      -0.63818    0.09091    -7.02    2.6e-12 ***
## Year2002      -0.72422    0.08946    -8.10    7.7e-16 ***
## Year2003      -0.73247    0.09023    -8.12    6.5e-16 ***
## Year2004      -0.66038    0.08508    -7.76    1.1e-14 ***
## Year2005      -0.71233    0.08578    -8.30    < 2e-16 ***
## Year2006      -0.69441    0.08476    -8.19    3.5e-16 ***
## Year2007      -0.78720    0.08373    -9.40    < 2e-16 ***
## Year2008      -0.73436    0.08461    -8.68    < 2e-16 ***
## Year2009      -0.69958    0.08399    -8.33    < 2e-16 ***
## Year2010      -0.72053    0.08383    -8.60    < 2e-16 ***
## Year2011      -0.79533    0.08321    -9.56    < 2e-16 ***
## Year2012      -0.75408    0.08438    -8.94    < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.152, Adjusted R-squared:  0.147
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 286 weights are ~= 1. The remaining 3354 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0081 0.8600 0.9500 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.75e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1          1.012
## LastAuthorFemale 1.028 1          1.014
## Year              1.051 16          1.002

```



## Residuals from first and last author



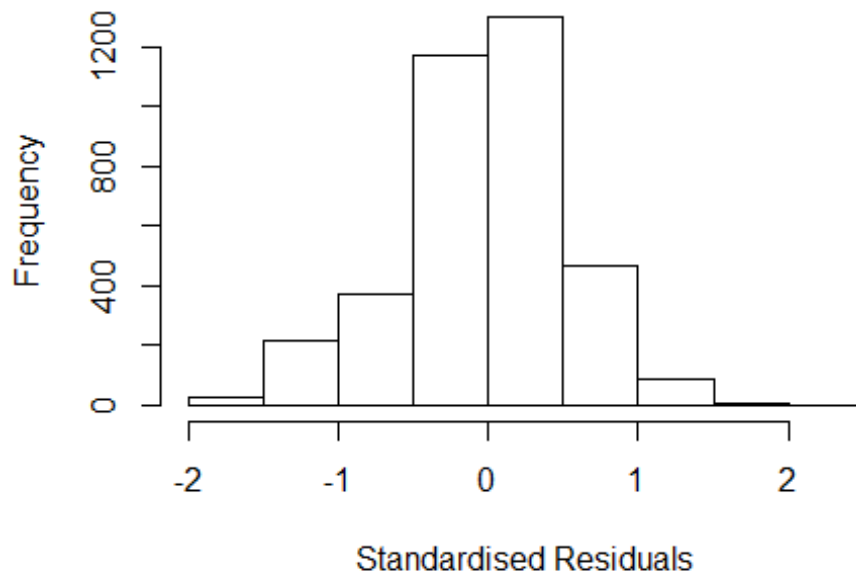
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8274 -0.3251  0.0161  0.3430  2.2844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.8137     0.0829   21.87 < 2e-16 ***
## FirstAuthorFemale1  0.0136     0.0187    0.73  0.466
## LastAuthorFemale1 -0.0506     0.0224   -2.27  0.024 *
## Year1997          -0.2709     0.1061   -2.55  0.011 *
## Year1998          -0.6751     0.0920   -7.34 2.6e-13 ***
## Year1999          -0.5857     0.0888   -6.59 4.9e-11 ***
## Year2000          -0.5814     0.0932   -6.24 4.9e-10 ***
## Year2001          -0.6321     0.0957   -6.61 4.5e-11 ***
## Year2002          -0.7104     0.0948   -7.50 8.2e-14 ***
## Year2003          -0.7092     0.0949   -7.48 9.6e-14 ***
## Year2004          -0.6530     0.0894   -7.30 3.4e-13 ***
## Year2005          -0.6822     0.0901   -7.57 4.7e-14 ***
```

```

## Year2006          -0.6611      0.0890    -7.43  1.3e-13 ***
## Year2007          -0.7570      0.0880    -8.60  < 2e-16 ***
## Year2008          -0.6993      0.0886    -7.89  4.0e-15 ***
## Year2009          -0.6497      0.0881    -7.37  2.1e-13 ***
## Year2010          -0.6929      0.0878    -7.89  4.0e-15 ***
## Year2011          -0.7319      0.0872    -8.39  < 2e-16 ***
## Year2012          -0.7108      0.0889    -7.99  1.7e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0863, Adjusted R-squared:  0.0818
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 319 weights are ~= 1. The remaining 3321 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8580 0.9490 0.8900 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year      1.025 16      1.001

```

## Residuals from first author



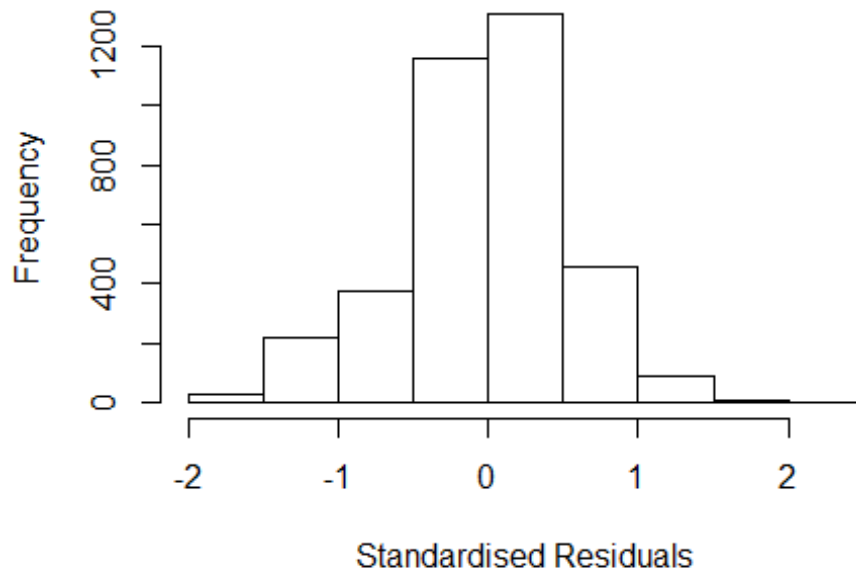
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8177 -0.3278 0.0136 0.3366 2.3003
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.80984 0.08327 21.73 < 2e-16 ***
## FirstAuthorFemale1 0.00785 0.01886 0.42 0.68
## Year1997 -0.27344 0.10643 -2.57 0.01 *
## Year1998 -0.67834 0.09223 -7.36 2.3e-13 ***
## Year1999 -0.58873 0.08911 -6.61 4.5e-11 ***
## Year2000 -0.58410 0.09357 -6.24 4.8e-10 ***
## Year2001 -0.63664 0.09603 -6.63 3.9e-11 ***
## Year2002 -0.71706 0.09498 -7.55 5.5e-14 ***
## Year2003 -0.71114 0.09508 -7.48 9.3e-14 ***
## Year2004 -0.65559 0.08972 -7.31 3.3e-13 ***
## Year2005 -0.68801 0.09025 -7.62 3.1e-14 ***
## Year2006 -0.66273 0.08925 -7.43 1.4e-13 ***
```

```

## Year2007          -0.76272      0.08822      -8.65 < 2e-16 ***
## Year2008          -0.70508      0.08888      -7.93 2.8e-15 ***
## Year2009          -0.65728      0.08827      -7.45 1.2e-13 ***
## Year2010          -0.69810      0.08807      -7.93 3.0e-15 ***
## Year2011          -0.73632      0.08742      -8.42 < 2e-16 ***
## Year2012          -0.71699      0.08905      -8.05 1.1e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0855, Adjusted R-squared:  0.0812
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 3497 is an outlier with |weight| <= 5.1e-07 ( < 2.7e-05);
## 325 weights are ~= 1. The remaining 3314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0548 0.8570 0.9480 0.8900 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.75e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.028 1          1.014
## Year          1.028 16          1.001

```

## Residuals from last author



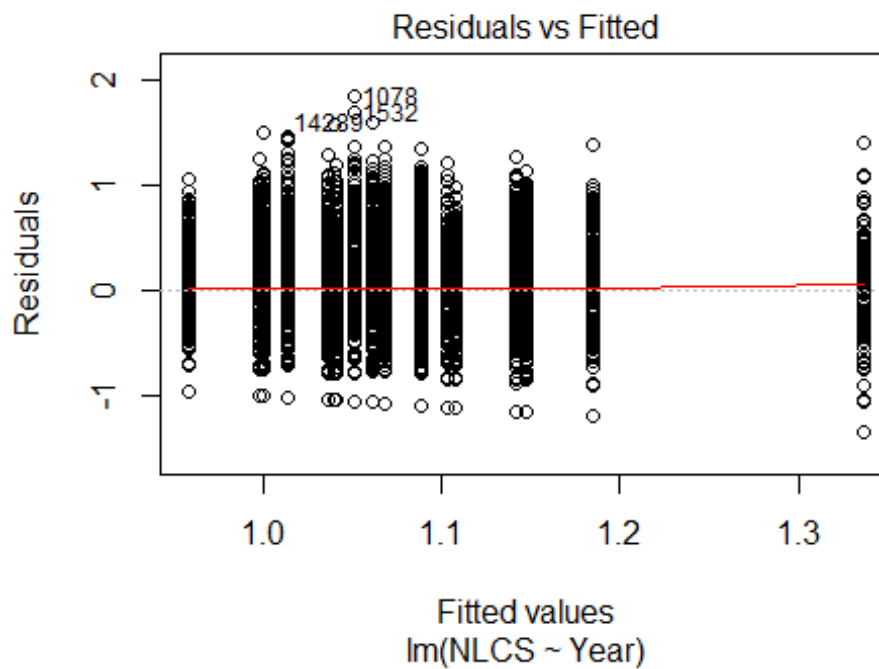
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8168 -0.3222 0.0156 0.3421 2.2941
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8168 0.0827 21.98 < 2e-16 ***
## LastAuthorFemale1 -0.0485 0.0225 -2.16 0.031 *
## Year1997 -0.2705 0.1059 -2.55 0.011 *
## Year1998 -0.6743 0.0919 -7.34 2.7e-13 ***
## Year1999 -0.5843 0.0887 -6.59 5.1e-11 ***
## Year2000 -0.5820 0.0931 -6.25 4.5e-10 ***
## Year2001 -0.6321 0.0956 -6.61 4.4e-11 ***
## Year2002 -0.7098 0.0947 -7.50 8.1e-14 ***
## Year2003 -0.7075 0.0947 -7.47 1.0e-13 ***
## Year2004 -0.6522 0.0893 -7.30 3.4e-13 ***
## Year2005 -0.6805 0.0900 -7.56 5.0e-14 ***
## Year2006 -0.6604 0.0888 -7.43 1.3e-13 ***
```

```

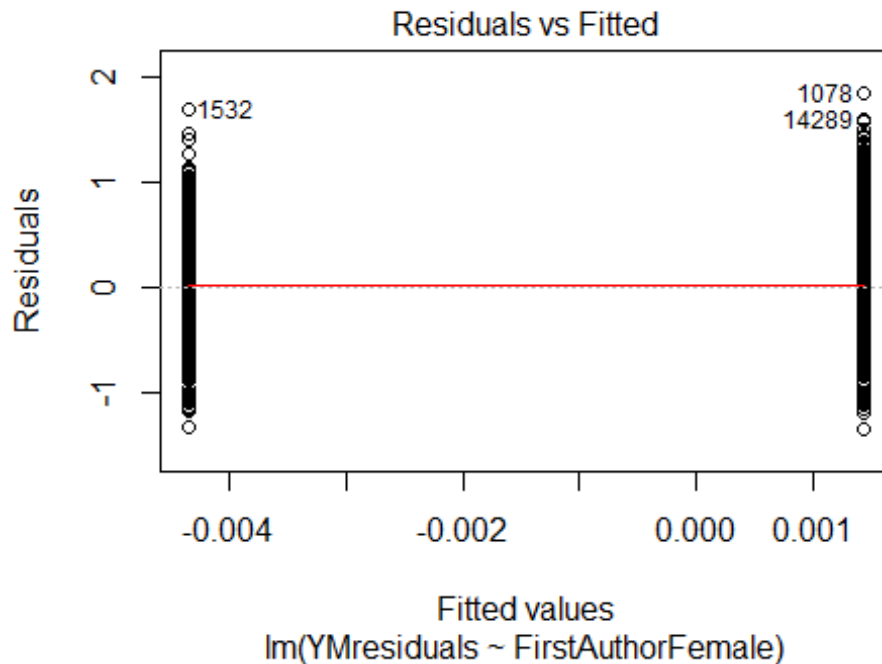
## Year2007          -0.7558      0.0879   -8.60 < 2e-16 ***
## Year2008          -0.6986      0.0885   -7.89 4.0e-15 ***
## Year2009          -0.6484      0.0880   -7.37 2.1e-13 ***
## Year2010          -0.6911      0.0877   -7.88 4.2e-15 ***
## Year2011          -0.7308      0.0871   -8.39 < 2e-16 ***
## Year2012          -0.7099      0.0888   -8.00 1.7e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.0862, Adjusted R-squared:  0.0819
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 316 weights are ~= 1. The remaining 3324 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000  0.859  0.948  0.890  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3640"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2732"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  470  477  557  540  648  612  634  603  684  817  858  954  931  988  889
## 2011 2012
##  947 1005
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  123  174  208  194  187  165  264  263  309  391  465  549  544  580  548
## 2011 2012

```

```
## 569 626
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 158 187 164 162 144 229 204 260 307 364 439 435 496 481
## 2011 2012
## 512 547
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 45, df = 16, p-value = 1e-04
```



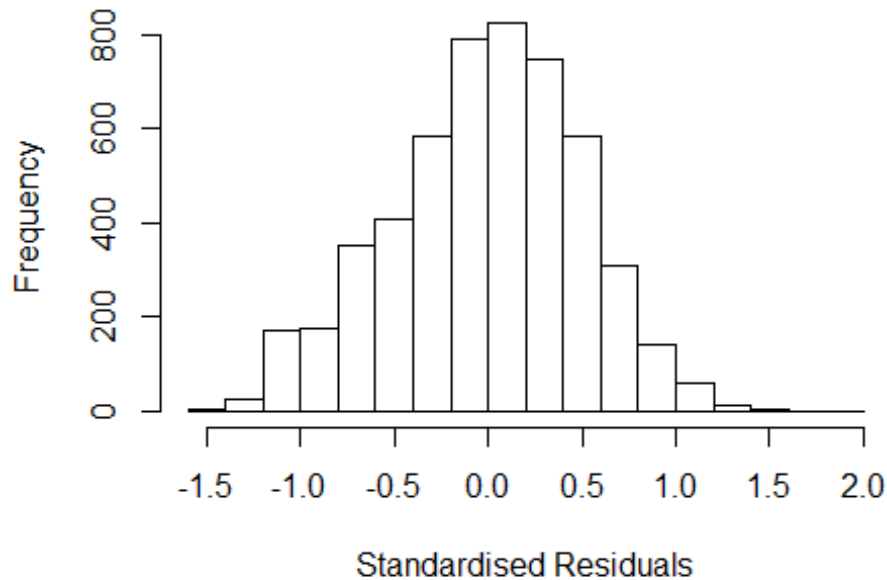
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.7, df = 1, p-value = 0.06
```



```
## [1] "Female first author team size 2018 geometric mean: 3.65339176159814"
## [1] "Male first author team size 2018 geometric mean: 3.49909300890247"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 45000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.17883397506202"
## [1] "Male last author team size 2018 geometric mean: 3.63812995407686"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.074 1 1.036
## LastAuthorFemale 1.071 1 1.035
## UniqueAuthors 1.116 4 1.014
## Year 1.142 16 1.004
```



## Residuals from first and last author and team size



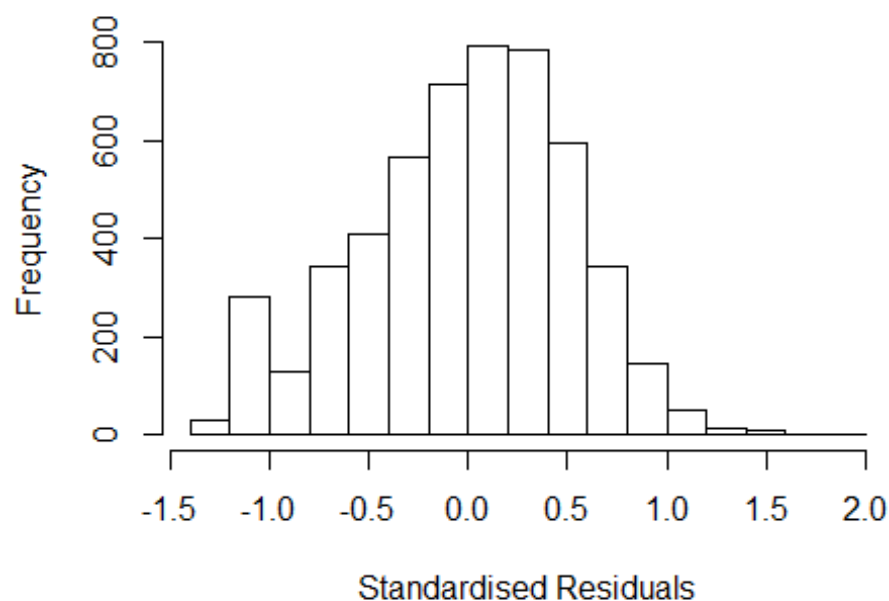
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4991 -0.3354  0.0198  0.3465  1.8138
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16826    0.05742   20.35 < 2e-16 ***
## FirstAuthorFemale1  0.00696    0.01697    0.41  0.682
## LastAuthorFemale1 -0.01887    0.01864   -1.01  0.312
## UniqueAuthors2     0.24623    0.02842    8.66 < 2e-16 ***
## UniqueAuthors3     0.30000    0.02766   10.84 < 2e-16 ***
## UniqueAuthors4     0.34118    0.02876   11.86 < 2e-16 ***
## UniqueAuthors5     0.46749    0.02729   17.13 < 2e-16 ***
## Year1997          -0.14356    0.06823   -2.10  0.035 *
## Year1998          -0.33732    0.07405   -4.56 5.4e-06 ***
## Year1999          -0.40072    0.07086   -5.66 1.6e-08 ***
```

```

## Year2000      -0.33745    0.06697    -5.04    4.8e-07 ***
## Year2001      -0.29943    0.06747    -4.44    9.3e-06 ***
## Year2002      -0.31286    0.06190    -5.05    4.5e-07 ***
## Year2003      -0.43056    0.06575    -6.55    6.4e-11 ***
## Year2004      -0.39270    0.06275    -6.26    4.2e-10 ***
## Year2005      -0.40912    0.06087    -6.72    2.0e-11 ***
## Year2006      -0.45918    0.06084    -7.55    5.2e-14 ***
## Year2007      -0.26974    0.05920    -4.56    5.3e-06 ***
## Year2008      -0.30842    0.05914    -5.21    1.9e-07 ***
## Year2009      -0.39918    0.05981    -6.67    2.7e-11 ***
## Year2010      -0.35883    0.06004    -5.98    2.4e-09 ***
## Year2011      -0.45224    0.05975    -7.57    4.4e-14 ***
## Year2012      -0.38730    0.05879    -6.59    4.9e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.503
## Multiple R-squared:  0.0855, Adjusted R-squared:  0.0816
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 440 weights are ~= 1. The remaining 4757 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.166  0.865   0.949   0.907   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.92e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.072 1          1.035
## LastAuthorFemale 1.067 1          1.033
## Year 1.052 16          1.002

```

## Residuals from first and last author



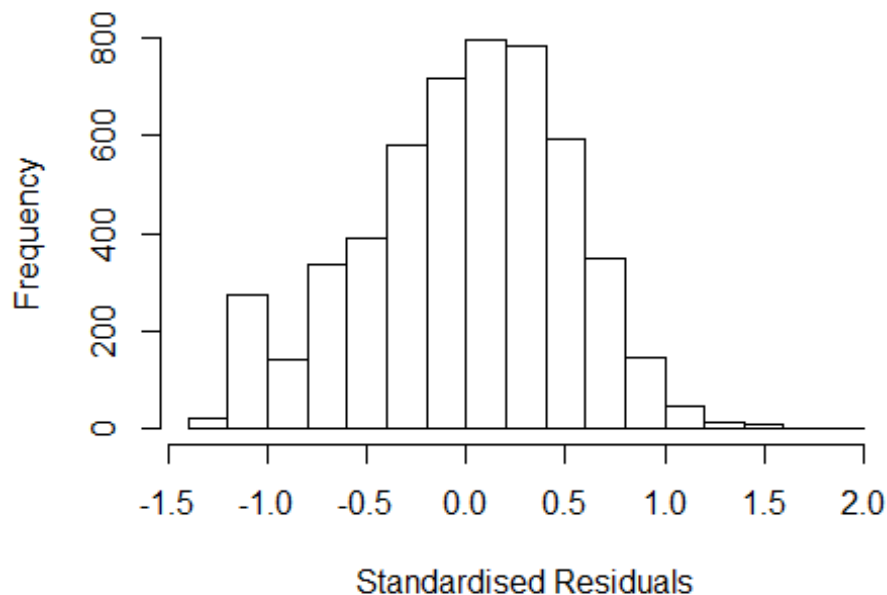
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.393 -0.349 0.029 0.358 1.829
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.39302 0.05627 24.75 < 2e-16 ***
## FirstAuthorFemale1 -0.00116 0.01775 -0.07 0.94782
## LastAuthorFemale1 -0.05015 0.01988 -2.52 0.01169 *
## Year1997 -0.17278 0.07152 -2.42 0.01572 *
## Year1998 -0.33054 0.07822 -4.23 2.4e-05 ***
## Year1999 -0.38527 0.07517 -5.13 3.1e-07 ***
## Year2000 -0.31140 0.06922 -4.50 7.0e-06 ***
## Year2001 -0.26049 0.07114 -3.66 0.00025 ***
## Year2002 -0.25884 0.06446 -4.02 6.0e-05 ***
## Year2003 -0.37028 0.06754 -5.48 4.4e-08 ***
## Year2004 -0.31912 0.06510 -4.90 9.8e-07 ***
## Year2005 -0.32684 0.06279 -5.21 2.0e-07 ***
```

```

## Year2006      -0.38986    0.06334   -6.15  8.1e-10 ***
## Year2007      -0.18112    0.06103   -2.97  0.00301 **
## Year2008      -0.20594    0.06121   -3.36  0.00077 ***
## Year2009      -0.30549    0.06166   -4.95  7.5e-07 ***
## Year2010      -0.26205    0.06194   -4.23  2.4e-05 ***
## Year2011      -0.35742    0.06167   -5.80  7.2e-09 ***
## Year2012      -0.28192    0.06078   -4.64  3.6e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.0185
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 420 weights are ~= 1. The remaining 4777 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.185  0.865  0.949  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1      1.019
## Year      1.038 16      1.001

```

## Residuals from first author



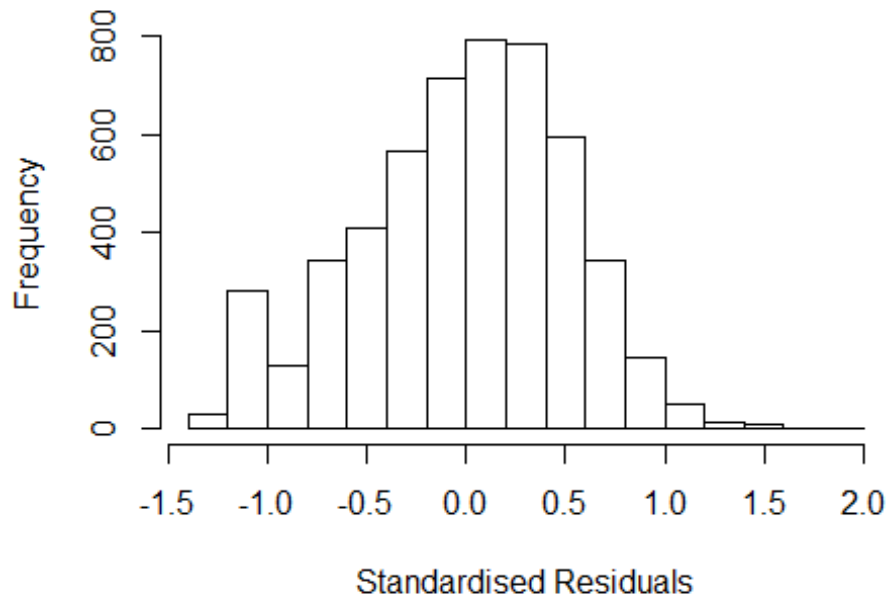
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3865 -0.3505 0.0288 0.3579 1.8361
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3865 0.0563 24.65 < 2e-16 ***
## FirstAuthorFemale1 -0.0154 0.0177 -0.87 0.38536
## Year1997 -0.1728 0.0718 -2.41 0.01610 *
## Year1998 -0.3316 0.0787 -4.21 2.5e-05 ***
## Year1999 -0.3837 0.0754 -5.09 3.7e-07 ***
## Year2000 -0.3133 0.0692 -4.52 6.2e-06 ***
## Year2001 -0.2639 0.0712 -3.71 0.00021 ***
## Year2002 -0.2574 0.0645 -3.99 6.8e-05 ***
## Year2003 -0.3692 0.0676 -5.46 4.9e-08 ***
## Year2004 -0.3179 0.0652 -4.88 1.1e-06 ***
## Year2005 -0.3246 0.0629 -5.16 2.5e-07 ***
## Year2006 -0.3904 0.0634 -6.16 7.7e-10 ***
```

```

## Year2007          -0.1801      0.0611   -2.95  0.00321 **
## Year2008          -0.2048      0.0613   -3.34  0.00083 ***
## Year2009          -0.3043      0.0617   -4.93  8.4e-07 ***
## Year2010          -0.2616      0.0620   -4.22  2.5e-05 ***
## Year2011          -0.3583      0.0617   -5.81  6.7e-09 ***
## Year2012          -0.2808      0.0609   -4.61  4.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0206, Adjusted R-squared:  0.0174
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 420 weights are ~= 1. The remaining 4777 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.181  0.866  0.949  0.906  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1          1.015
## Year            1.029 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3929 -0.3489 0.0286 0.3581 1.8288
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3929 0.0563 24.74 < 2e-16 ***
## LastAuthorFemale1 -0.0506 0.0197 -2.57 0.01011 *
## Year1997 -0.1728 0.0715 -2.42 0.01567 *
## Year1998 -0.3307 0.0782 -4.23 2.4e-05 ***
## Year1999 -0.3853 0.0752 -5.13 3.1e-07 ***
## Year2000 -0.3115 0.0692 -4.50 6.8e-06 ***
## Year2001 -0.2605 0.0711 -3.66 0.00025 ***
## Year2002 -0.2589 0.0644 -4.02 5.9e-05 ***
## Year2003 -0.3704 0.0674 -5.49 4.1e-08 ***
## Year2004 -0.3192 0.0650 -4.91 9.2e-07 ***
## Year2005 -0.3269 0.0627 -5.21 1.9e-07 ***
## Year2006 -0.3899 0.0633 -6.16 7.8e-10 ***
```

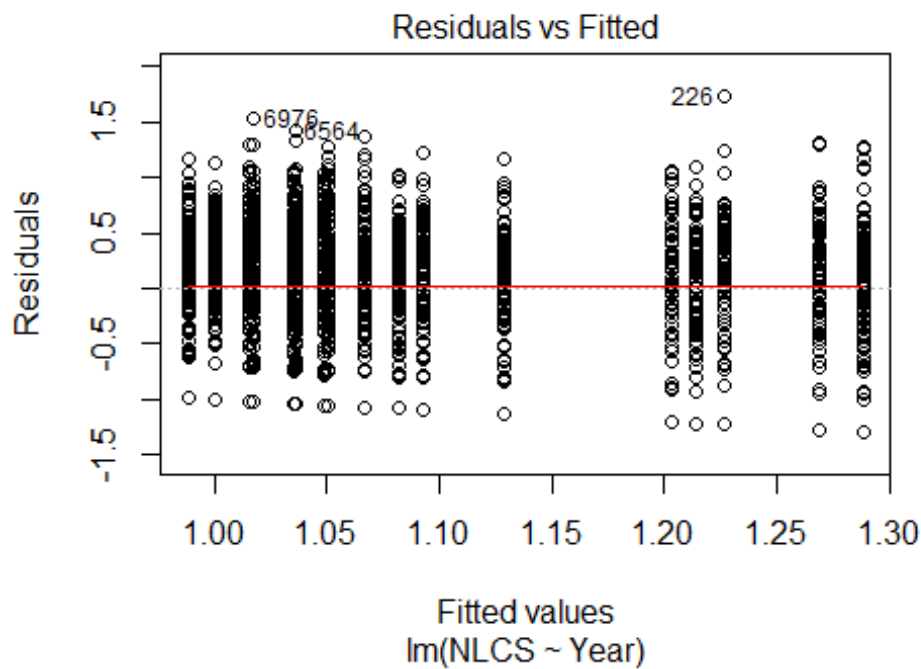
```

## Year2007          -0.1812      0.0610   -2.97  0.00297 **
## Year2008          -0.2061      0.0611   -3.37  0.00075 ***
## Year2009          -0.3056      0.0616   -4.96  7.1e-07 ***
## Year2010          -0.2622      0.0618   -4.24  2.3e-05 ***
## Year2011          -0.3575      0.0616   -5.80  7.0e-09 ***
## Year2012          -0.2820      0.0607   -4.65  3.5e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.0219, Adjusted R-squared:  0.0187
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 418 weights are ~= 1. The remaining 4779 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.185  0.865  0.949  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5197"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2733"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 473 381 331 435 472 456 429 375 384 394 420 451 406 416 373
## 2011 2012
## 384 353
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 90 99 79 112 96 113 158 167 147 170 167 168 158 165 181
## 2011 2012

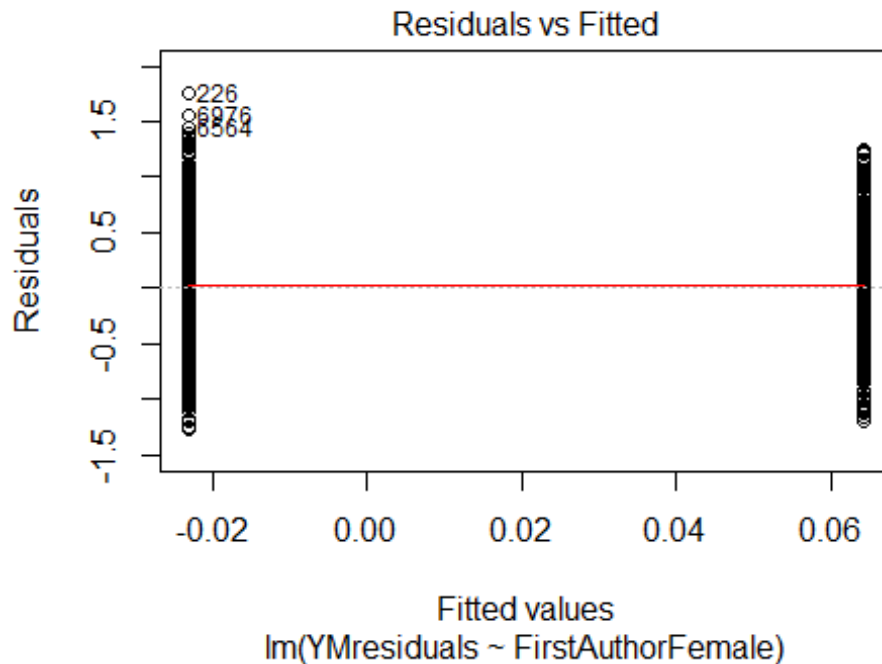
```



```
## 189 161
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 84 85 67 81 77 80 119 94 94 112 109 144 138 145 156
## 2011 2012
## 155 134
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.06
```

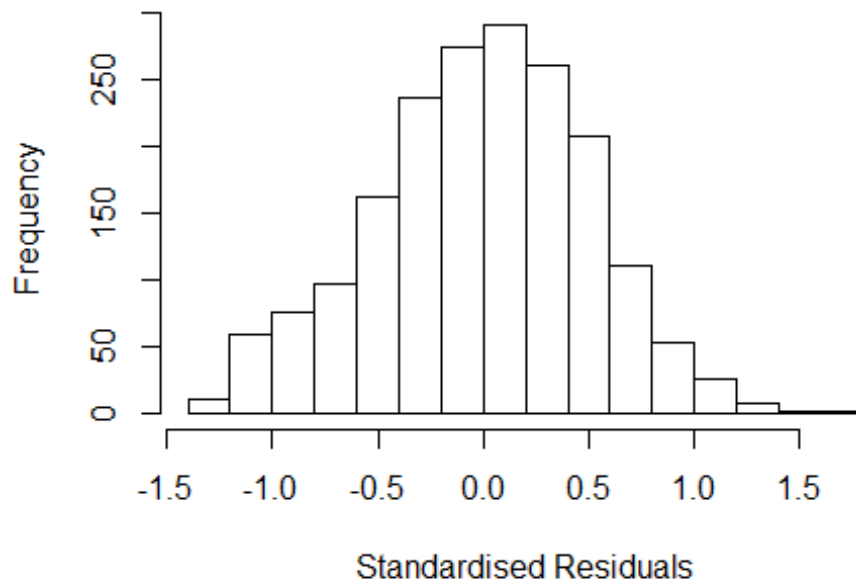


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.62209998809839"
## [1] "Male first author team size 2018 geometric mean: 3.81377730805822"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 880, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.33045950098654"
## [1] "Male last author team size 2018 geometric mean: 3.89159423354974"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 670, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.098 1      1.048
## LastAuthorFemale  1.134 1      1.065
## UniqueAuthors    1.187 4      1.022
## Year             1.257 16      1.007
```

## Residuals from first and last author and team size



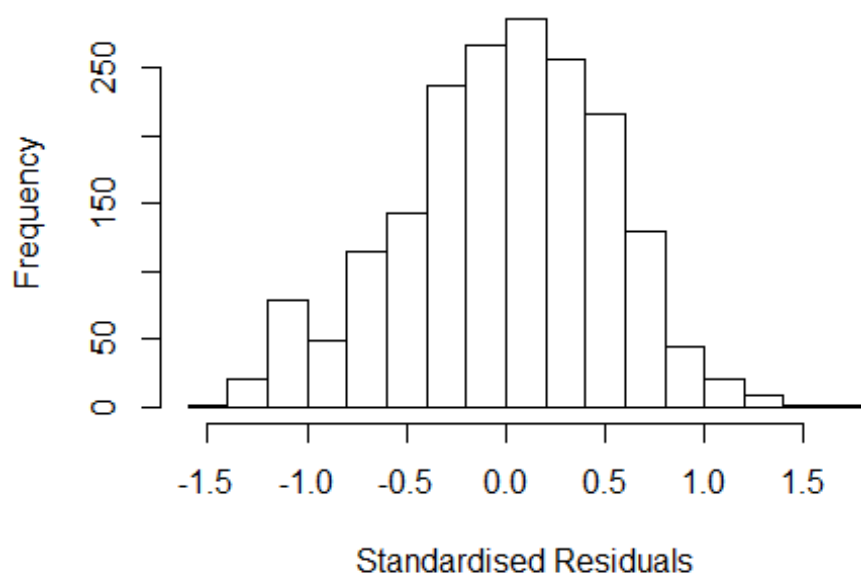
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3462 -0.3298 0.0204 0.3456 1.7411
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0392 0.0840 12.38 < 2e-16 ***
## FirstAuthorFemale1 0.0634 0.0271 2.34 0.01940 *
## LastAuthorFemale1 0.1617 0.0315 5.13 3.1e-07 ***
## UniqueAuthors2 0.1195 0.0572 2.09 0.03691 *
## UniqueAuthors3 0.1767 0.0550 3.21 0.00135 **
## UniqueAuthors4 0.1985 0.0565 3.51 0.00046 ***
## UniqueAuthors5 0.4027 0.0559 7.20 8.6e-13 ***
## Year1997 0.0572 0.0868 0.66 0.50957
## Year1998 0.0311 0.1033 0.30 0.76366
## Year1999 0.0258 0.0866 0.30 0.76559
```

```

## Year2000          -0.0536      0.0871   -0.61  0.53877
## Year2001          -0.1531      0.0906   -1.69  0.09118 .
## Year2002          -0.1512      0.0801   -1.89  0.05911 .
## Year2003          -0.1880      0.0810   -2.32  0.02038 *
## Year2004          -0.1934      0.0948   -2.04  0.04147 *
## Year2005          -0.2236      0.0803   -2.78  0.00543 **
## Year2006          -0.2280      0.0838   -2.72  0.00658 **
## Year2007          -0.2107      0.0798   -2.64  0.00835 **
## Year2008          -0.2796      0.0781   -3.58  0.00035 ***
## Year2009          -0.2181      0.0808   -2.70  0.00704 **
## Year2010          -0.2593      0.0788   -3.29  0.00102 **
## Year2011          -0.2350      0.0786   -2.99  0.00283 **
## Year2012          -0.2931      0.0810   -3.62  0.00030 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0914, Adjusted R-squared:  0.0806
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1736 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.217  0.875   0.952   0.909   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.34e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1          1.043
## LastAuthorFemale  1.121 1          1.059
## Year              1.073 16          1.002

```

## Residuals from first and last author



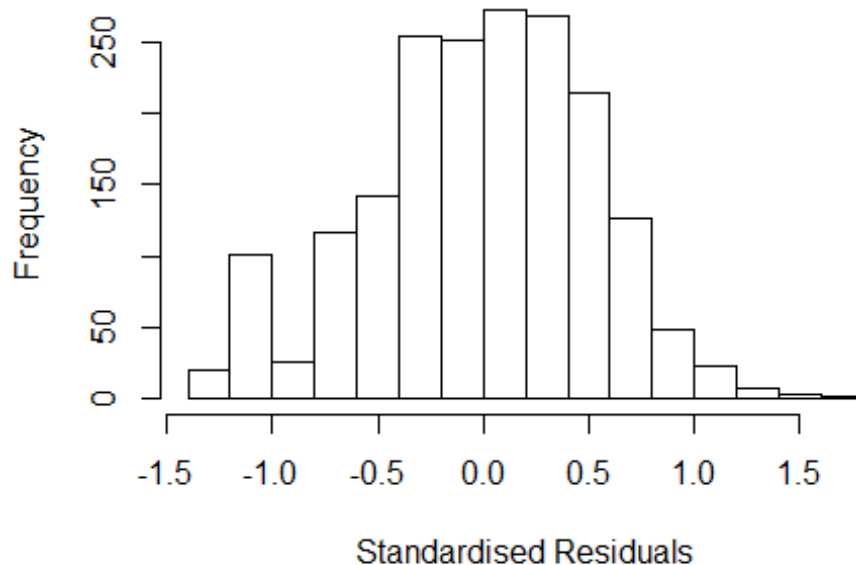
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4081 -0.3381 0.0238 0.3650 1.7512
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2058 0.0703 17.14 <2e-16 ***
## FirstAuthorFemale1 0.0623 0.0278 2.24 0.0251 *
## LastAuthorFemale1 0.1579 0.0320 4.93 9e-07 ***
## Year1997 0.0576 0.0904 0.64 0.5242
## Year1998 0.0125 0.1049 0.12 0.9049
## Year1999 0.0444 0.0884 0.50 0.6157
## Year2000 -0.0351 0.0894 -0.39 0.6949
## Year2001 -0.1358 0.0940 -1.45 0.1486
## Year2002 -0.1118 0.0839 -1.33 0.1830
## Year2003 -0.1642 0.0846 -1.94 0.0524 .
## Year2004 -0.1666 0.0967 -1.72 0.0853 .
## Year2005 -0.1951 0.0841 -2.32 0.0205 *
```

```

## Year2006          -0.2120      0.0861   -2.46   0.0139 *
## Year2007          -0.1733      0.0831   -2.09   0.0372 *
## Year2008          -0.2191      0.0809   -2.71   0.0068 **
## Year2009          -0.1852      0.0836   -2.21   0.0269 *
## Year2010          -0.2097      0.0817   -2.57   0.0103 *
## Year2011          -0.1876      0.0817   -2.30   0.0217 *
## Year2012          -0.2238      0.0845   -2.65   0.0082 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0451, Adjusted R-squared:  0.0358
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 140 weights are ~= 1. The remaining 1734 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.234  0.872  0.952  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.014 1      1.007
## Year      1.014 16      1.000

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2802 -0.3451 0.0243 0.3637 1.7379
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2191 0.0711 17.14 < 2e-16 ***
## FirstAuthorFemale1 0.0974 0.0271 3.60 0.00033 ***
## Year1997 0.0611 0.0909 0.67 0.50164
## Year1998 0.0239 0.1038 0.23 0.81791
## Year1999 0.0458 0.0894 0.51 0.60856
## Year2000 -0.0276 0.0897 -0.31 0.75857
## Year2001 -0.1317 0.0946 -1.39 0.16404
## Year2002 -0.1015 0.0847 -1.20 0.23061
## Year2003 -0.1594 0.0861 -1.85 0.06441 .
## Year2004 -0.1649 0.0982 -1.68 0.09346 .
## Year2005 -0.1884 0.0852 -2.21 0.02705 *
## Year2006 -0.1866 0.0872 -2.14 0.03252 *
```

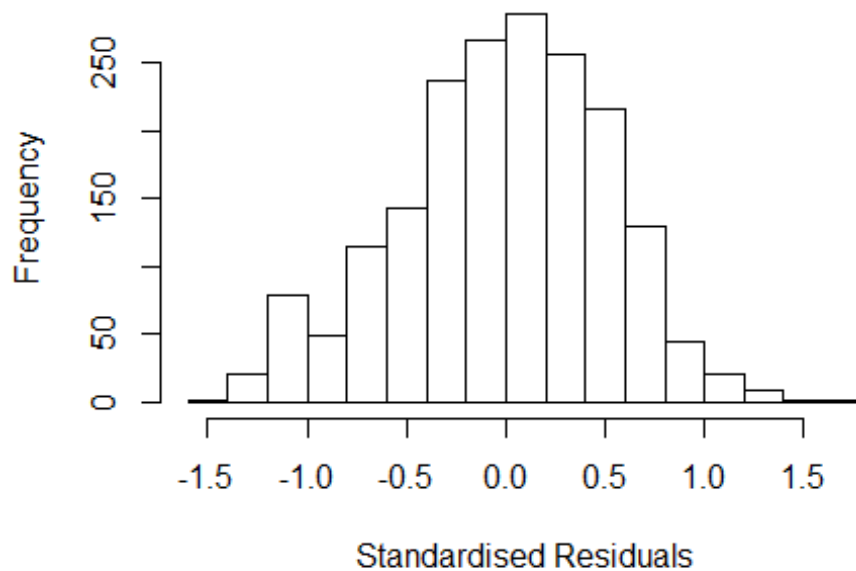
```

## Year2007          -0.1634      0.0846   -1.93  0.05367 .
## Year2008          -0.2117      0.0819   -2.59  0.00979 **
## Year2009          -0.1789      0.0841   -2.13  0.03356 *
## Year2010          -0.2071      0.0827   -2.50  0.01234 *
## Year2011          -0.1939      0.0827   -2.35  0.01911 *
## Year2012          -0.2121      0.0851   -2.49  0.01279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0326, Adjusted R-squared:  0.0238
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1733 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.248  0.874  0.951  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.051 1          1.025
## Year            1.051 16          1.002

```



## Residuals from last author



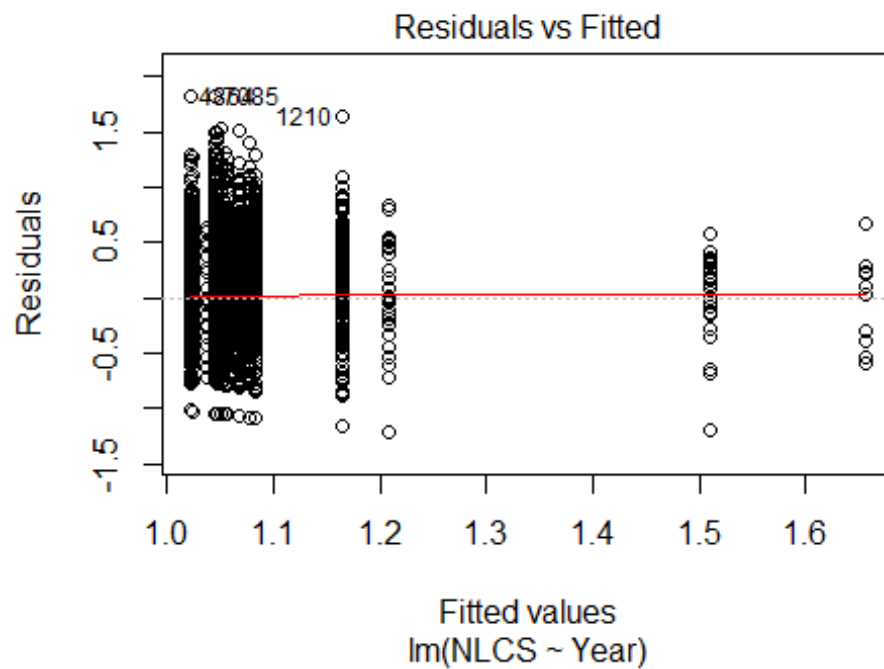
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4420 -0.3480  0.0264  0.3628  1.7395
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2175     0.0701   17.36 < 2e-16 ***
## LastAuthorFemale1  0.1765     0.0310    5.69 1.5e-08 ***
## Year1997          0.0603     0.0911    0.66  0.5084
## Year1998          0.0107     0.1051    0.10  0.9186
## Year1999          0.0481     0.0886    0.54  0.5870
## Year2000         -0.0327     0.0897   -0.36  0.7153
## Year2001         -0.1340     0.0940   -1.43  0.1543
## Year2002         -0.1095     0.0843   -1.30  0.1940
## Year2003         -0.1601     0.0850   -1.88  0.0598 .
## Year2004         -0.1662     0.0967   -1.72  0.0858 .
## Year2005         -0.1945     0.0843   -2.31  0.0212 *
## Year2006         -0.2088     0.0862   -2.42  0.0155 *
```

```

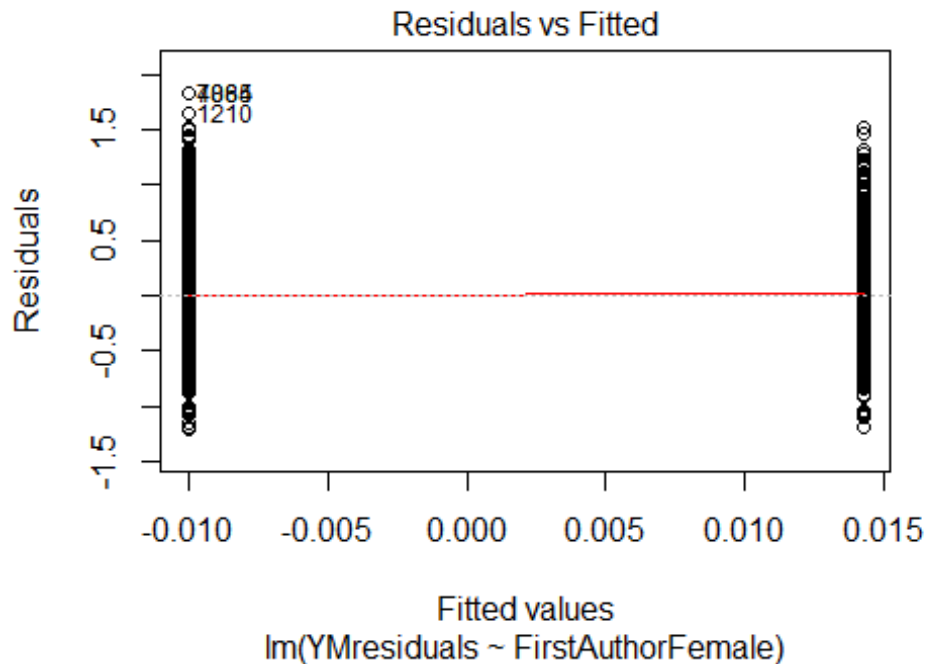
## Year2007          -0.1684      0.0832   -2.03   0.0429 *
## Year2008          -0.2174      0.0812   -2.68   0.0074 **
## Year2009          -0.1812      0.0840   -2.16   0.0311 *
## Year2010          -0.2081      0.0819   -2.54   0.0111 *
## Year2011          -0.1790      0.0817   -2.19   0.0287 *
## Year2012          -0.2226      0.0845   -2.63   0.0085 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.52
## Multiple R-squared:  0.0426, Adjusted R-squared:  0.0338
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1725 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.240  0.872  0.951  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1874"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2734"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  536  512  41   31   56  463   55  371  348  340  403  407  456  488  454
## 2011 2012
##  484  467
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  171  208   24   11   27  146   23  175  162  178  218  250  268  316  307
## 2011 2012

```

```
## 362 347
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 157 193 23 10 25 134 20 152 138 150 190 217 232 282 272
## 2011 2012
## 320 312
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```

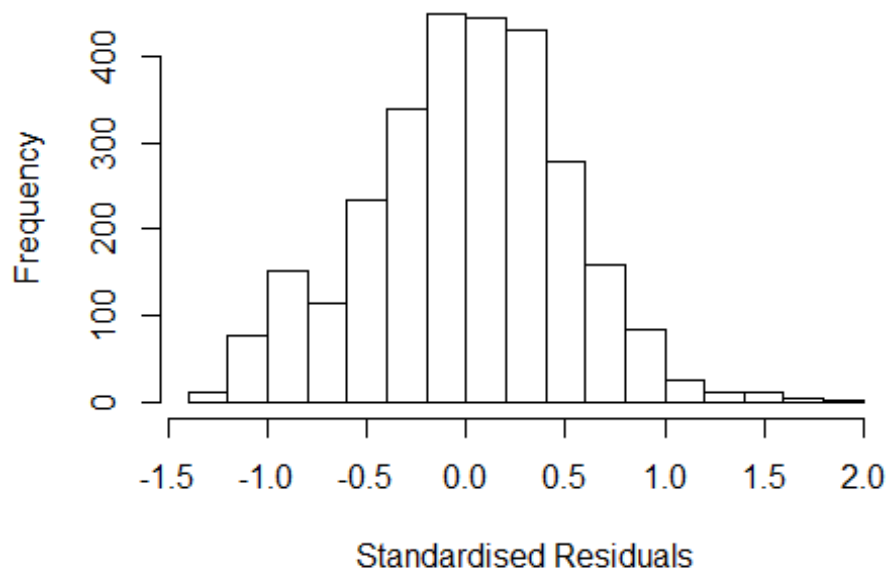


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 2.72949359703148"
## [1] "Male first author team size 2018 geometric mean: 2.69334671299988"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.40896151409882"
## [1] "Male last author team size 2018 geometric mean: 2.9634130233778"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8300, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.092 1          1.045
## LastAuthorFemale  1.118 1          1.057
## UniqueAuthors    1.154 4          1.018
## Year              1.188 16         1.005
```

## Residuals from first and last author and team size



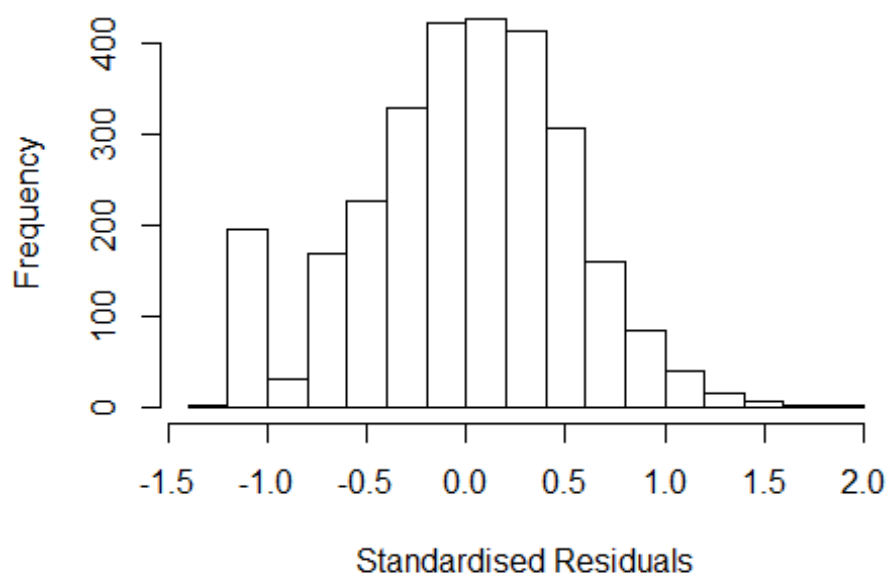
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2878 -0.3339 0.0184 0.3336 1.8531
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.81892 0.04524 18.10 < 2e-16 ***
## FirstAuthorFemale1 0.00403 0.02023 0.20 0.842
## LastAuthorFemale1 -0.01356 0.02183 -0.62 0.535
## UniqueAuthors2 0.20977 0.03018 6.95 4.5e-12 ***
## UniqueAuthors3 0.17483 0.03298 5.30 1.2e-07 ***
## UniqueAuthors4 0.30585 0.03506 8.72 < 2e-16 ***
## UniqueAuthors5 0.43379 0.02818 15.39 < 2e-16 ***
## Year1997 0.12698 0.05471 2.32 0.020 *
## Year1998 0.55378 0.11269 4.91 9.4e-07 ***
## Year1999 0.69956 0.13585 5.15 2.8e-07 ***
```

```

## Year2000          0.21189    0.09209    2.30    0.021 *
## Year2001          0.07014    0.05650    1.24    0.215
## Year2002          0.02868    0.10977    0.26    0.794
## Year2003         -0.00156    0.06215   -0.03    0.980
## Year2004          0.03861    0.06368    0.61    0.544
## Year2005          0.00997    0.05682    0.18    0.861
## Year2006          0.02716    0.05946    0.46    0.648
## Year2007          0.03415    0.05261    0.65    0.516
## Year2008          0.01559    0.05438    0.29    0.774
## Year2009          0.03292    0.05277    0.62    0.533
## Year2010          0.03021    0.05103    0.59    0.554
## Year2011          0.03101    0.05082    0.61    0.542
## Year2012          0.00905    0.05558    0.16    0.871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.49
## Multiple R-squared:  0.108, Adjusted R-squared:  0.101
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 232 weights are ~= 1. The remaining 2595 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.122  0.866  0.950  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.54e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1          1.043
## LastAuthorFemale  1.093 1          1.045
## Year              1.079 16          1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.241 -0.342 0.024 0.356 1.801
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0081 0.0500 20.15 < 2e-16 ***
## FirstAuthorFemale1 0.0384 0.0213 1.80 0.071 .
## LastAuthorFemale1 -0.0443 0.0225 -1.97 0.049 *
## Year1997 0.1548 0.0627 2.47 0.014 *
## Year1998 0.5399 0.0938 5.75 9.7e-09 ***
## Year1999 0.6752 0.1380 4.89 1.0e-06 ***
## Year2000 0.2333 0.1096 2.13 0.033 *
## Year2001 0.0846 0.0653 1.29 0.196
## Year2002 0.0694 0.1099 0.63 0.528
## Year2003 0.0700 0.0711 0.98 0.325
## Year2004 0.0805 0.0715 1.13 0.260
## Year2005 0.0418 0.0651 0.64 0.521
```

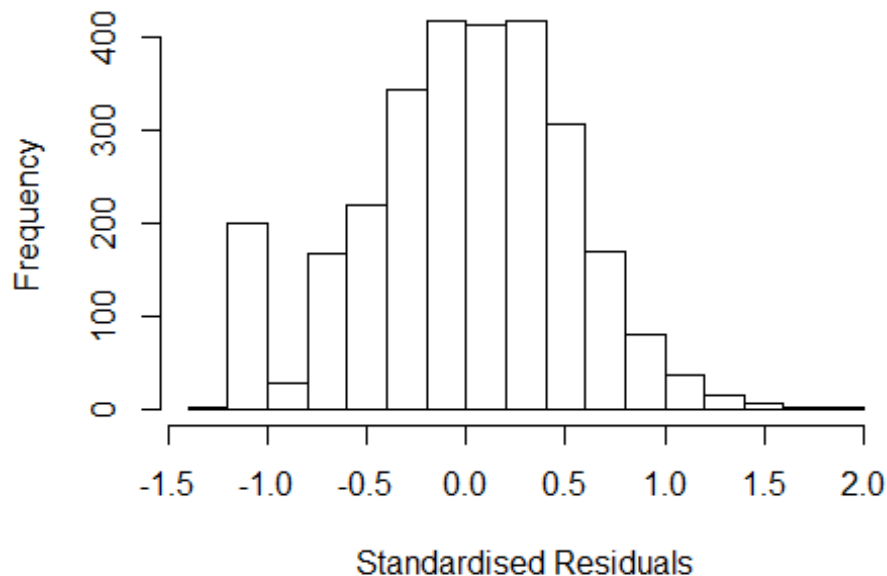
```

## Year2006          0.0322      0.0665      0.48      0.628
## Year2007          0.0419      0.0589      0.71      0.477
## Year2008          0.0306      0.0604      0.51      0.612
## Year2009          0.0721      0.0592      1.22      0.223
## Year2010          0.0516      0.0585      0.88      0.379
## Year2011          0.0583      0.0582      1.00      0.316
## Year2012          0.0538      0.0623      0.86      0.388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.0121
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 216 weights are ~= 1. The remaining 2611 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.192  0.867  0.950  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```



## Residuals from first author



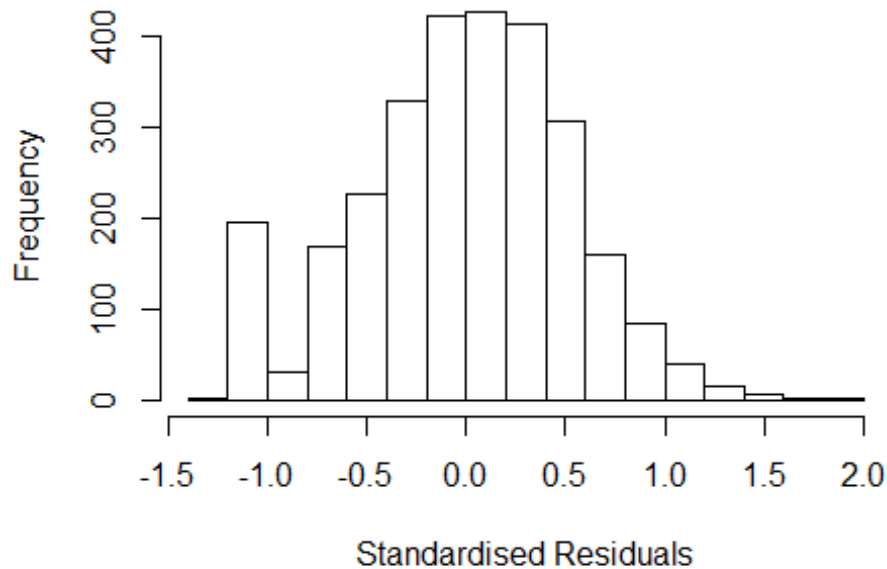
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2286 -0.3375  0.0249  0.3499  1.8150
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0028     0.0500   20.04 < 2e-16 ***
## FirstAuthorFemale1 0.0274     0.0209    1.31  0.189
## Year1997          0.1536     0.0630    2.44  0.015 *
## Year1998          0.5367     0.0946    5.67 1.5e-08 ***
## Year1999          0.6778     0.1413    4.80 1.7e-06 ***
## Year2000          0.2258     0.1087    2.08  0.038 *
## Year2001          0.0820     0.0654    1.25  0.210
## Year2002          0.0634     0.1105    0.57  0.566
## Year2003          0.0662     0.0712    0.93  0.353
## Year2004          0.0748     0.0713    1.05  0.295
## Year2005          0.0366     0.0651    0.56  0.574
## Year2006          0.0247     0.0664    0.37  0.709
```

```

## Year2007          0.0363      0.0589      0.62      0.537
## Year2008          0.0259      0.0605      0.43      0.668
## Year2009          0.0668      0.0592      1.13      0.259
## Year2010          0.0483      0.0587      0.82      0.411
## Year2011          0.0526      0.0582      0.90      0.367
## Year2012          0.0452      0.0622      0.73      0.467
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.513
## Multiple R-squared:  0.017, Adjusted R-squared:  0.0111
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 219 weights are ~= 1. The remaining 2608 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.185  0.868  0.951   0.903  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.046 1      1.023
## Year              1.046 16      1.001

```

## Residuals from last author



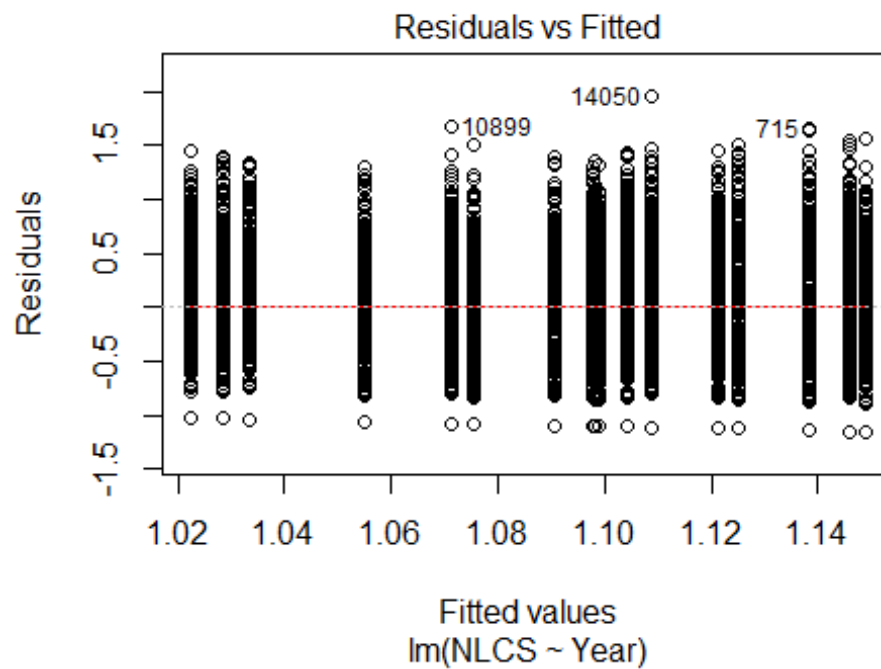
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2487 -0.3439 0.0271 0.3543 1.7871
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0155 0.0498 20.41 < 2e-16 ***
## LastAuthorFemale1 -0.0339 0.0220 -1.54 0.123
## Year1997 0.1573 0.0629 2.50 0.013 *
## Year1998 0.5384 0.0934 5.76 9.1e-09 ***
## Year1999 0.6839 0.1353 5.05 4.6e-07 ***
## Year2000 0.2332 0.1090 2.14 0.033 *
## Year2001 0.0898 0.0653 1.38 0.169
## Year2002 0.0679 0.1096 0.62 0.536
## Year2003 0.0773 0.0711 1.09 0.277
## Year2004 0.0862 0.0713 1.21 0.227
## Year2005 0.0486 0.0649 0.75 0.454
## Year2006 0.0402 0.0663 0.61 0.544
```

```

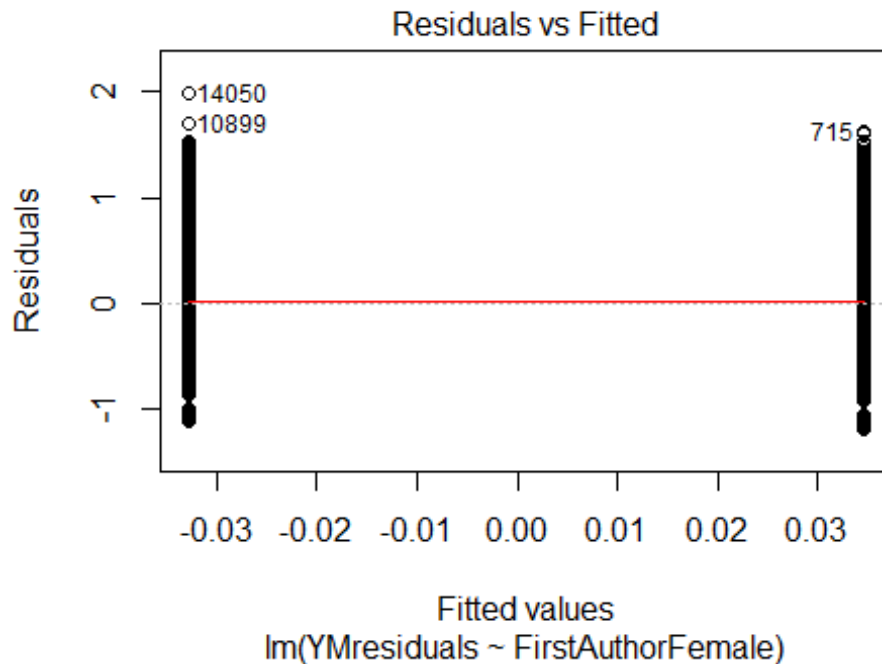
## Year2007          0.0471      0.0587      0.80      0.422
## Year2008          0.0368      0.0603      0.61      0.542
## Year2009          0.0773      0.0591      1.31      0.191
## Year2010          0.0554      0.0584      0.95      0.343
## Year2011          0.0656      0.0578      1.13      0.257
## Year2012          0.0604      0.0621      0.97      0.330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.0172, Adjusted R-squared:  0.0113
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 205 weights are ~= 1. The remaining 2622 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.202  0.868  0.951  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2827"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2735"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 664 731 761 699 778 823 751 667 700 731 816 841 822 789 776
## 2011 2012
## 748 719
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 238 249 294 307 335 247 367 319 350 418 522 578 508 495 531
## 2011 2012

```

```
## 537 524
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 221 228 271 276 294 212 304 252 278 357 452 499 432 415 435
## 2011 2012
## 475 462
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 71, df = 16, p-value = 5e-09
```

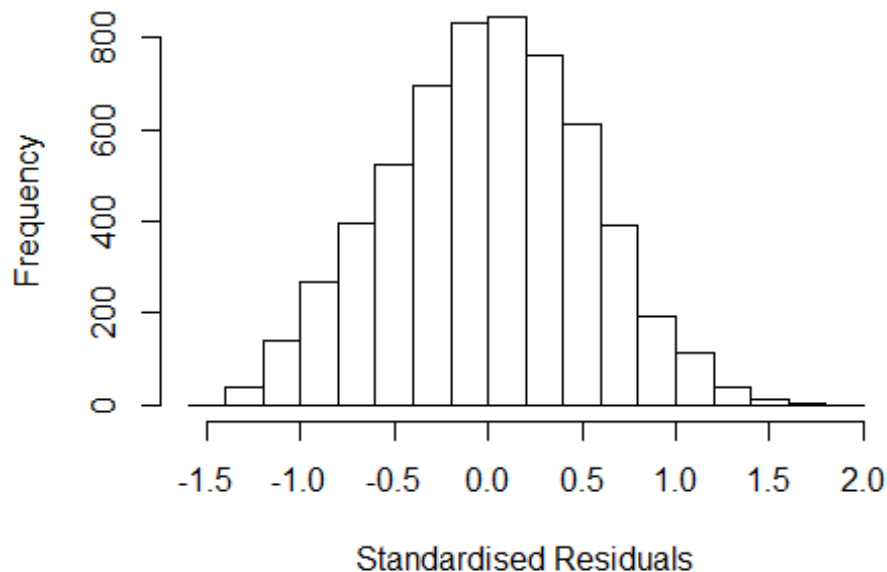


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.8, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 4.15408978022882"
## [1] "Male first author team size 2018 geometric mean: 3.45688652231836"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 27000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.85617381757547"
## [1] "Male last author team size 2018 geometric mean: 3.86981178511788"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1          1.045
## LastAuthorFemale  1.104 1          1.051
## UniqueAuthors    1.133 4          1.016
## Year              1.125 16         1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40685 -0.36658 0.00949 0.36899 1.88855
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8721 0.0498 17.51 < 2e-16 ***
## FirstAuthorFemale1 0.0556 0.0151 3.68 0.00023 ***
## LastAuthorFemale1 0.0407 0.0158 2.57 0.01020 *
## UniqueAuthors2 0.2420 0.0260 9.30 < 2e-16 ***
## UniqueAuthors3 0.2958 0.0260 11.38 < 2e-16 ***
## UniqueAuthors4 0.4074 0.0265 15.39 < 2e-16 ***
## UniqueAuthors5 0.5417 0.0239 22.65 < 2e-16 ***
## Year1997 -0.0625 0.0595 -1.05 0.29367
## Year1998 -0.0885 0.0594 -1.49 0.13578
## Year1999 -0.0704 0.0574 -1.23 0.21946
```

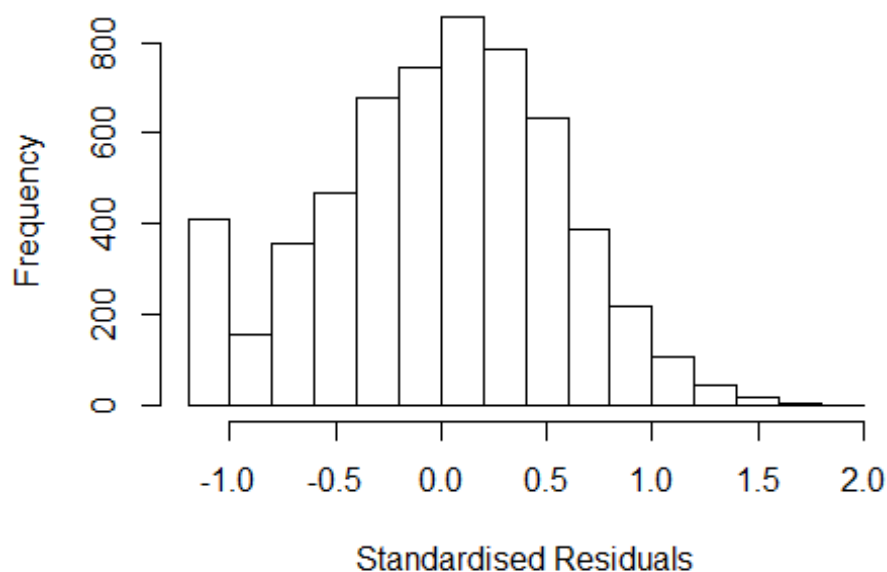
```

## Year2000          -0.0276      0.0552   -0.50   0.61706
## Year2001          -0.1407      0.0595   -2.36   0.01813 *
## Year2002          -0.1916      0.0558   -3.43   0.00060 ***
## Year2003          -0.1591      0.0563   -2.83   0.00469 **
## Year2004          -0.1264      0.0560   -2.26   0.02391 *
## Year2005          -0.1415      0.0536   -2.64   0.00832 **
## Year2006          -0.2127      0.0521   -4.09   4.4e-05 ***
## Year2007          -0.1288      0.0524   -2.46   0.01400 *
## Year2008          -0.1715      0.0532   -3.22   0.00127 **
## Year2009          -0.2148      0.0550   -3.90   9.6e-05 ***
## Year2010          -0.1166      0.0527   -2.21   0.02685 *
## Year2011          -0.1367      0.0536   -2.55   0.01076 *
## Year2012          -0.1689      0.0542   -3.11   0.00185 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.112, Adjusted R-squared:  0.108
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 467 weights are ~= 1. The remaining 5396 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.872  0.951   0.912   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.068 1      1.034
## LastAuthorFemale  1.066 1      1.033
## Year              1.017 16      1.001

```



## Residuals from first and last author



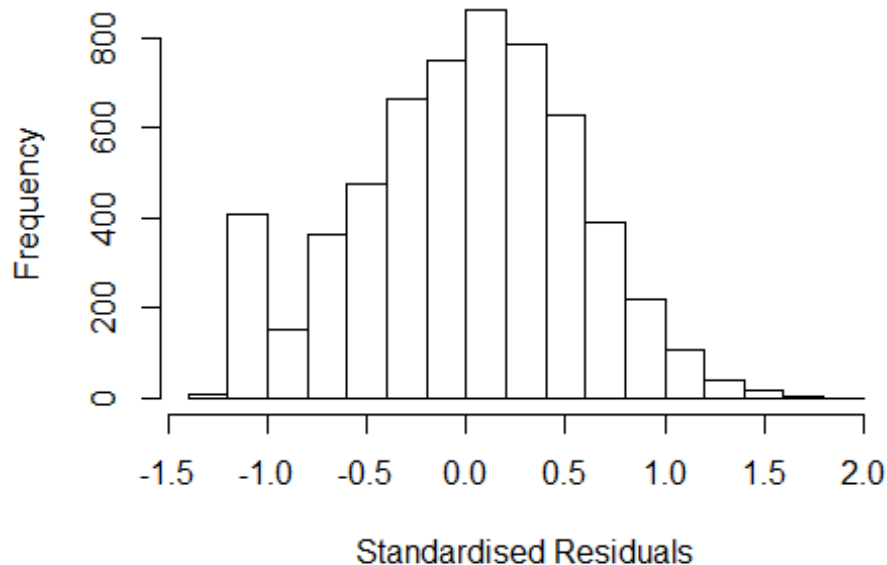
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1995 -0.3784 0.0312 0.3848 1.9889
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12131 0.05383 20.83 < 2e-16 ***
## FirstAuthorFemale1 0.07434 0.01582 4.70 2.7e-06 ***
## LastAuthorFemale1 -0.00926 0.01647 -0.56 0.574
## Year1997 -0.05670 0.06780 -0.84 0.403
## Year1998 -0.06610 0.06723 -0.98 0.326
## Year1999 -0.05125 0.06428 -0.80 0.425
## Year2000 0.01312 0.06168 0.21 0.832
## Year2001 -0.09501 0.06865 -1.38 0.166
## Year2002 -0.14434 0.06300 -2.29 0.022 *
## Year2003 -0.07172 0.06378 -1.12 0.261
## Year2004 -0.04388 0.06234 -0.70 0.482
## Year2005 -0.04809 0.05988 -0.80 0.422
```

```

## Year2006      -0.14203    0.05924   -2.40    0.017 *
## Year2007      -0.02868    0.05866   -0.49    0.625
## Year2008      -0.08756    0.05958   -1.47    0.142
## Year2009      -0.11553    0.06156   -1.88    0.061 .
## Year2010      -0.01080    0.05955   -0.18    0.856
## Year2011      -0.00193    0.05951   -0.03    0.974
## Year2012      -0.03896    0.06105   -0.64    0.523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00777
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 457 weights are ~= 1. The remaining 5406 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.189  0.870  0.950  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.01 1      1.005
## Year      1.01 16      1.000

```

## Residuals from first author



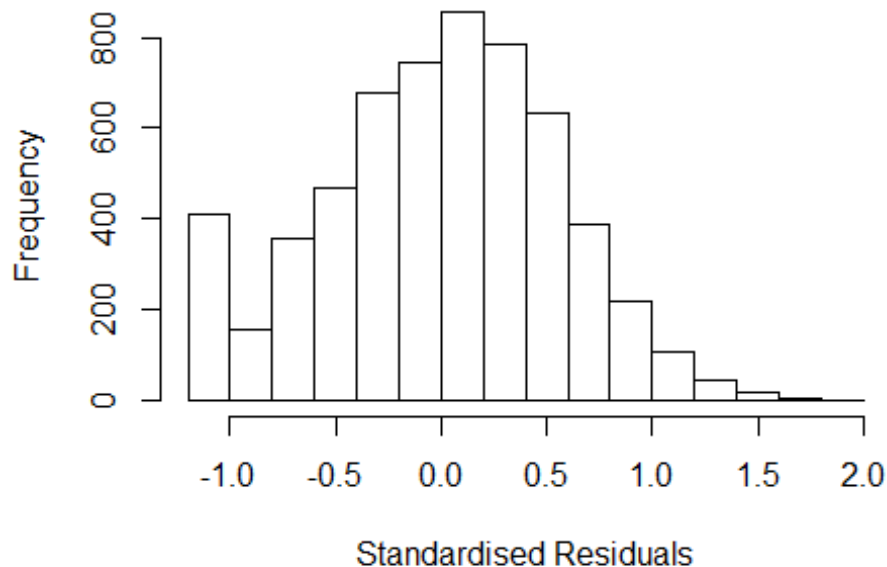
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2041 -0.3788 0.0301 0.3836 1.9817
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11898 0.05366 20.85 < 2e-16 ***
## FirstAuthorFemale1 0.07176 0.01541 4.66 3.3e-06 ***
## Year1997 -0.05688 0.06784 -0.84 0.402
## Year1998 -0.06569 0.06726 -0.98 0.329
## Year1999 -0.05097 0.06433 -0.79 0.428
## Year2000 0.01339 0.06171 0.22 0.828
## Year2001 -0.09511 0.06868 -1.38 0.166
## Year2002 -0.14385 0.06301 -2.28 0.022 *
## Year2003 -0.07143 0.06378 -1.12 0.263
## Year2004 -0.04380 0.06237 -0.70 0.483
## Year2005 -0.04788 0.05991 -0.80 0.424
## Year2006 -0.14185 0.05926 -2.39 0.017 *
```

```

## Year2007          -0.02828    0.05869   -0.48    0.630
## Year2008          -0.08705    0.05959   -1.46    0.144
## Year2009          -0.11523    0.06159   -1.87    0.061 .
## Year2010          -0.01021    0.05956   -0.17    0.864
## Year2011          -0.00164    0.05955   -0.03    0.978
## Year2012          -0.03865    0.06109   -0.63    0.527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.564
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00789
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 463 weights are ~= 1. The remaining 5400 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.192  0.869  0.950  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.008 1          1.004
## Year              1.008 16          1.000

```

## Residuals from last author



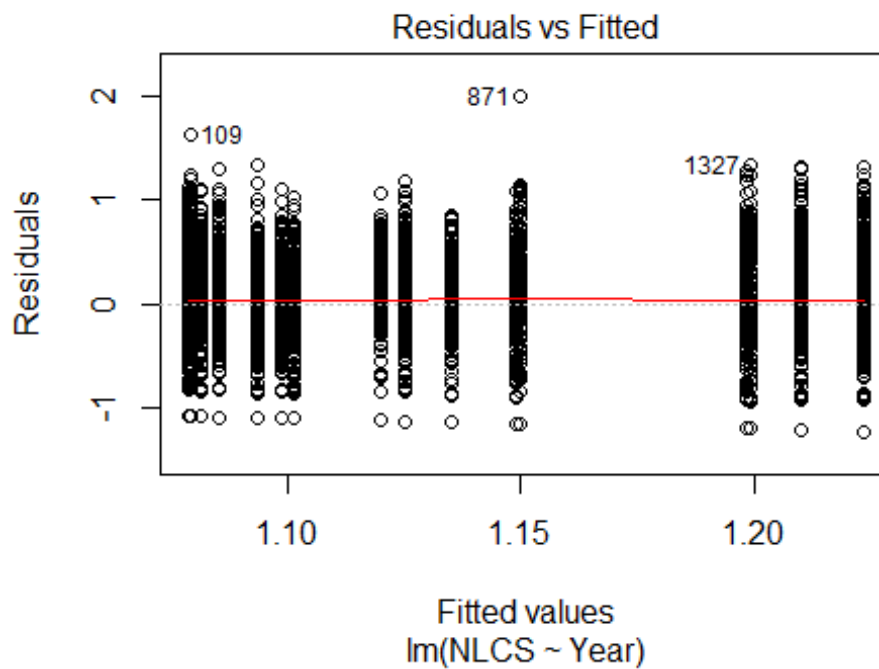
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1734 -0.3816 0.0262 0.3848 1.9344
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14509 0.05404 21.19 <2e-16 ***
## LastAuthorFemale1 0.01320 0.01602 0.82 0.410
## Year1997 -0.05301 0.06813 -0.78 0.437
## Year1998 -0.06323 0.06749 -0.94 0.349
## Year1999 -0.04810 0.06450 -0.75 0.456
## Year2000 0.01512 0.06216 0.24 0.808
## Year2001 -0.09021 0.06912 -1.31 0.192
## Year2002 -0.14050 0.06349 -2.21 0.027 *
## Year2003 -0.07078 0.06429 -1.10 0.271
## Year2004 -0.03975 0.06269 -0.63 0.526
## Year2005 -0.04546 0.06035 -0.75 0.451
## Year2006 -0.13802 0.05969 -2.31 0.021 *
```

```

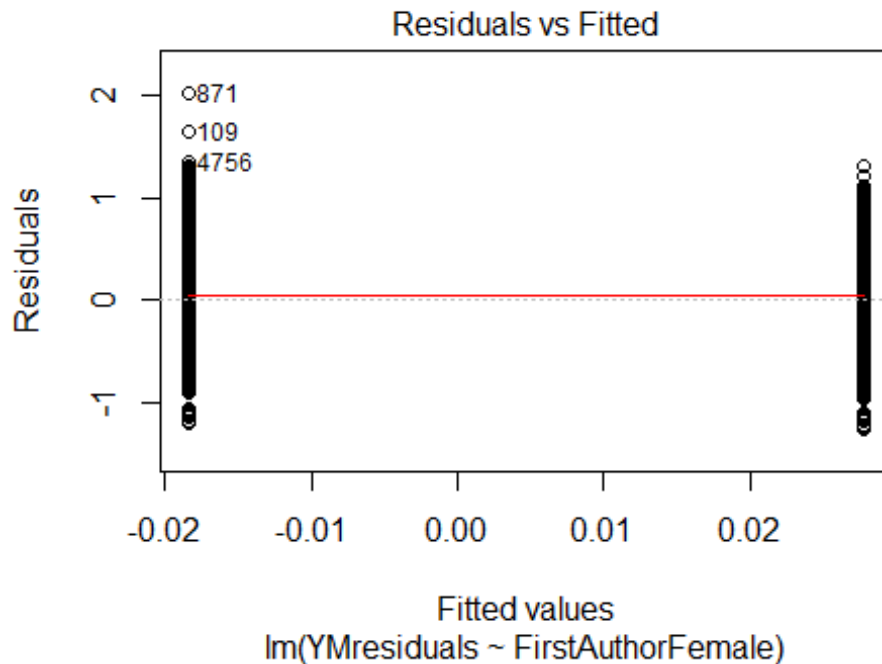
## Year2007          -0.02320      0.05904    -0.39      0.694
## Year2008          -0.08168      0.06001    -1.36      0.174
## Year2009          -0.10980      0.06195    -1.77      0.076 .
## Year2010          -0.00414      0.05994    -0.07      0.945
## Year2011           0.00679      0.05997      0.11      0.910
## Year2012          -0.03072      0.06149    -0.50      0.617
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.567
## Multiple R-squared:  0.00697,    Adjusted R-squared:  0.00408
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 451 weights are ~= 1. The remaining 5412 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.220  0.868  0.951  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5863"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2736"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  443  412  349  403  488  478  440  367  414  370  411  482  547  518  507
## 2011 2012
##  582  605
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  146  134  137  167  179  135  199  191  234  217  219  274  311  295  279
## 2011 2012

```

```
## 368 385
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 129 120 122 138 168 125 178 148 196 175 177 211 260 249 236
## 2011 2012
## 328 336
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 110, df = 16, p-value = 2e-15
```



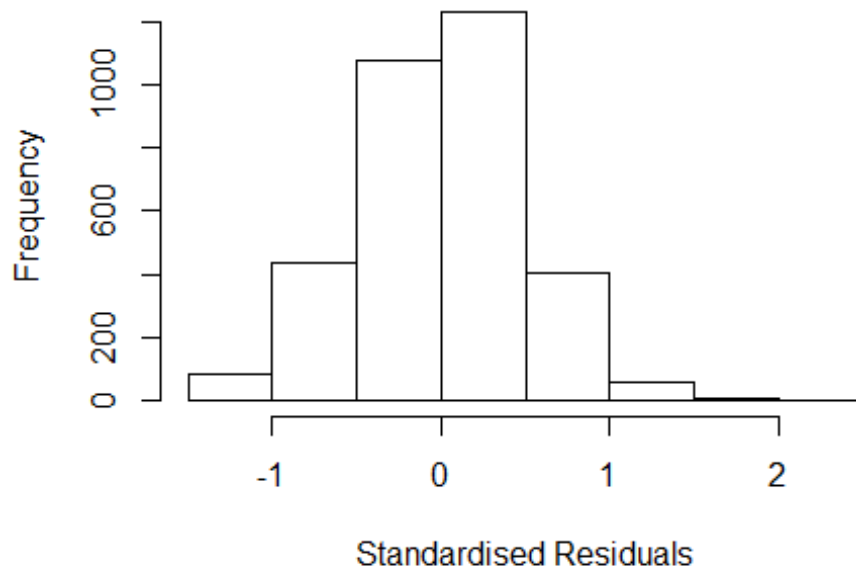
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 27, df = 1, p-value = 2e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 5.4986167823274"
## [1] "Male first author team size 2018 geometric mean: 4.70255349323402"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.29511263770434"
## [1] "Male last author team size 2018 geometric mean: 4.99307879733272"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 15000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061 1          1.030
## LastAuthorFemale  1.034 1          1.017
## UniqueAuthors    1.224 4          1.026
## Year             1.221 16          1.006
```



## Residuals from first and last author and team size



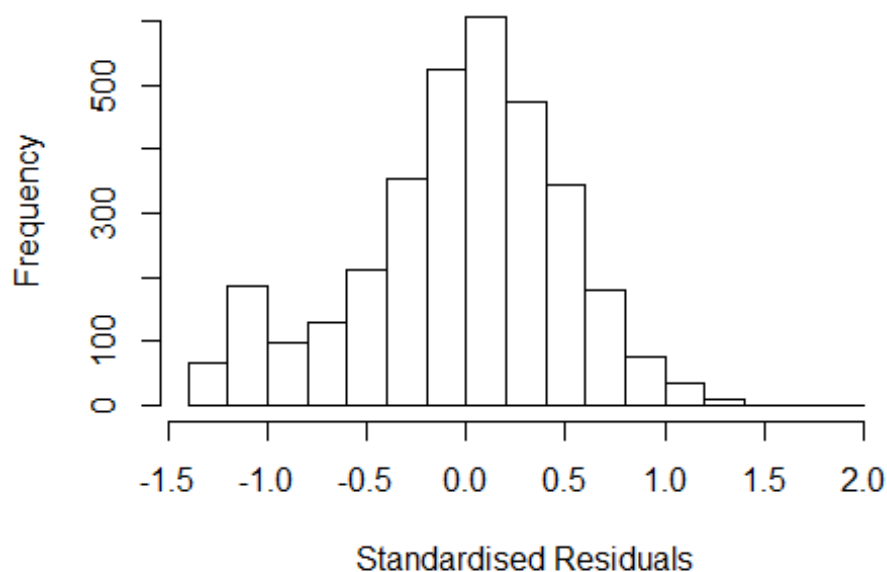
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3988 -0.3078 0.0202 0.3127 2.3826
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.69075 0.07330 9.42 < 2e-16 ***
## FirstAuthorFemale1 -0.00781 0.01775 -0.44 0.660
## LastAuthorFemale1 -0.01135 0.02141 -0.53 0.596
## UniqueAuthors2 0.31835 0.04371 7.28 4.1e-13 ***
## UniqueAuthors3 0.50855 0.04027 12.63 < 2e-16 ***
## UniqueAuthors4 0.52969 0.04107 12.90 < 2e-16 ***
## UniqueAuthors5 0.56811 0.03750 15.15 < 2e-16 ***
## Year1997 0.07260 0.09270 0.78 0.434
## Year1998 0.14780 0.08471 1.74 0.081 .
## Year1999 0.01851 0.08399 0.22 0.826
```

```

## Year2000      0.01528    0.07901    0.19    0.847
## Year2001      0.05493    0.08036    0.68    0.494
## Year2002      0.03809    0.07597    0.50    0.616
## Year2003     -0.00287    0.07494   -0.04    0.969
## Year2004     -0.04062    0.07424   -0.55    0.584
## Year2005     -0.05879    0.07592   -0.77    0.439
## Year2006     -0.03031    0.07636   -0.40    0.691
## Year2007     -0.01210    0.07564   -0.16    0.873
## Year2008     -0.00646    0.07577   -0.09    0.932
## Year2009      0.06774    0.07499    0.90    0.366
## Year2010     -0.01897    0.07551   -0.25    0.802
## Year2011      0.09454    0.07390    1.28    0.201
## Year2012      0.09643    0.07396    1.30    0.192
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.157, Adjusted R-squared:  0.151
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## observation 207 is an outlier with |weight| = 0 ( < 3e-05);
## 272 weights are ~= 1. The remaining 3023 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.107  0.851  0.948  0.897  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.03e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1 1.012
## LastAuthorFemale 1.030 1 1.015
## Year 1.050 16 1.002

```

## Residuals from first and last author



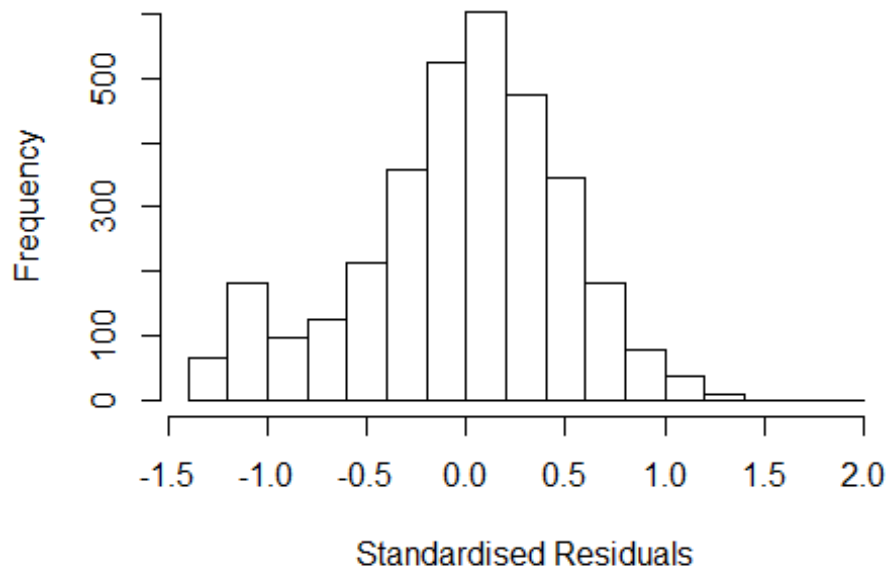
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.286 -0.314 0.028 0.319 1.963
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04873 0.07422 14.13 <2e-16 ***
## FirstAuthorFemale1 0.05273 0.01834 2.87 0.0041 **
## LastAuthorFemale1 -0.00344 0.02251 -0.15 0.8785
## Year1997 0.13426 0.10663 1.26 0.2081
## Year1998 0.15465 0.09150 1.69 0.0911 .
## Year1999 0.02582 0.09234 0.28 0.7798
## Year2000 0.03876 0.08544 0.45 0.6501
## Year2001 0.08248 0.08694 0.95 0.3429
## Year2002 0.10727 0.08132 1.32 0.1872
## Year2003 0.07217 0.08082 0.89 0.3719
## Year2004 0.02864 0.07965 0.36 0.7192
## Year2005 0.02473 0.08249 0.30 0.7643
```

```

## Year2006          0.03869    0.08284    0.47    0.6405
## Year2007          0.03613    0.07962    0.45    0.6500
## Year2008          0.03294    0.08145    0.40    0.6860
## Year2009          0.15391    0.07893    1.95    0.0513 .
## Year2010          0.08459    0.08101    1.04    0.2965
## Year2011          0.18024    0.07882    2.29    0.0223 *
## Year2012          0.18453    0.07920    2.33    0.0199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.018, Adjusted R-squared:  0.0126
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 270 weights are ~= 1. The remaining 3026 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0353 0.8490 0.9490 0.8860 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## Year              1.022 16      1.001

```

## Residuals from first author



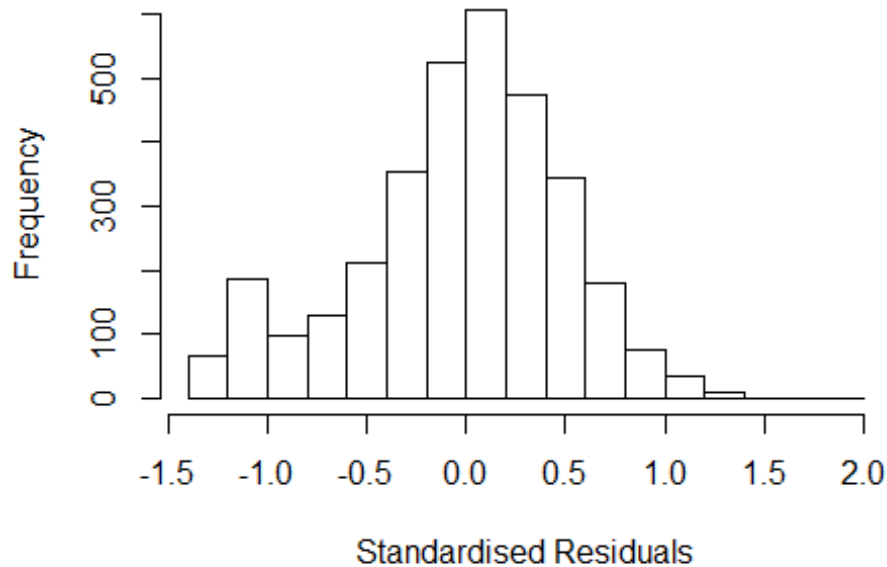
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2848 -0.3143 0.0279 0.3198 1.9635
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0483 0.0741 14.14 <2e-16 ***
## FirstAuthorFemale1 0.0524 0.0184 2.85 0.0044 **
## Year1997 0.1342 0.1066 1.26 0.2083
## Year1998 0.1547 0.0915 1.69 0.0911 .
## Year1999 0.0256 0.0924 0.28 0.7818
## Year2000 0.0386 0.0854 0.45 0.6515
## Year2001 0.0825 0.0869 0.95 0.3430
## Year2002 0.1073 0.0813 1.32 0.1869
## Year2003 0.0721 0.0808 0.89 0.3723
## Year2004 0.0285 0.0797 0.36 0.7205
## Year2005 0.0245 0.0825 0.30 0.7663
## Year2006 0.0383 0.0828 0.46 0.6434
```

```

## Year2007          0.0357      0.0796      0.45      0.6541
## Year2008          0.0326      0.0815      0.40      0.6887
## Year2009          0.1536      0.0790      1.95      0.0518 .
## Year2010          0.0843      0.0810      1.04      0.2983
## Year2011          0.1799      0.0789      2.28      0.0226 *
## Year2012          0.1841      0.0792      2.33      0.0201 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.018, Adjusted R-squared:  0.0129
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 269 weights are ~= 1. The remaining 3027 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0353 0.8480 0.9490 0.8860 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.029 1      1.014
## Year      1.029 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2592 -0.3098 0.0273 0.3212 1.9442
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06309 0.07451 14.27 <2e-16 ***
## LastAuthorFemale1 0.00437 0.02241 0.20 0.845
## Year1997 0.13873 0.10843 1.28 0.201
## Year1998 0.15551 0.09210 1.69 0.091 .
## Year1999 0.02578 0.09298 0.28 0.782
## Year2000 0.04334 0.08596 0.50 0.614
## Year2001 0.08823 0.08766 1.01 0.314
## Year2002 0.11082 0.08180 1.35 0.176
## Year2003 0.07778 0.08145 0.95 0.340
## Year2004 0.03280 0.08037 0.41 0.683
## Year2005 0.02972 0.08317 0.36 0.721
## Year2006 0.04406 0.08359 0.53 0.598
```

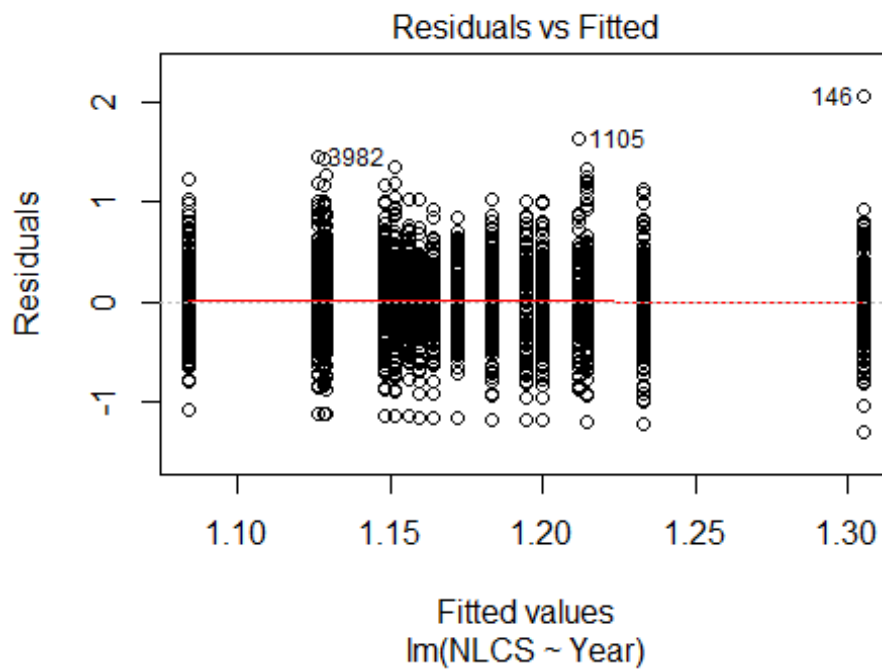
```

## Year2007      0.04519      0.08028      0.56      0.574
## Year2008      0.04045      0.08214      0.49      0.622
## Year2009      0.16170      0.07956      2.03      0.042 *
## Year2010      0.09330      0.08166      1.14      0.253
## Year2011      0.18549      0.07942      2.34      0.020 *
## Year2012      0.19172      0.07989      2.40      0.016 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.0155, Adjusted R-squared:  0.0104
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 269 weights are ~= 1. The remaining 3027 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0413 0.8480 0.9490 0.8860 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3296"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2737"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 292 267 241 218 254 256 279 202 212 223 270 275 295 351 318
## 2011 2012
## 321 328
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 149 140 107 120 117 133 160 120 127 129 193 181 207 268 224
## 2011 2012

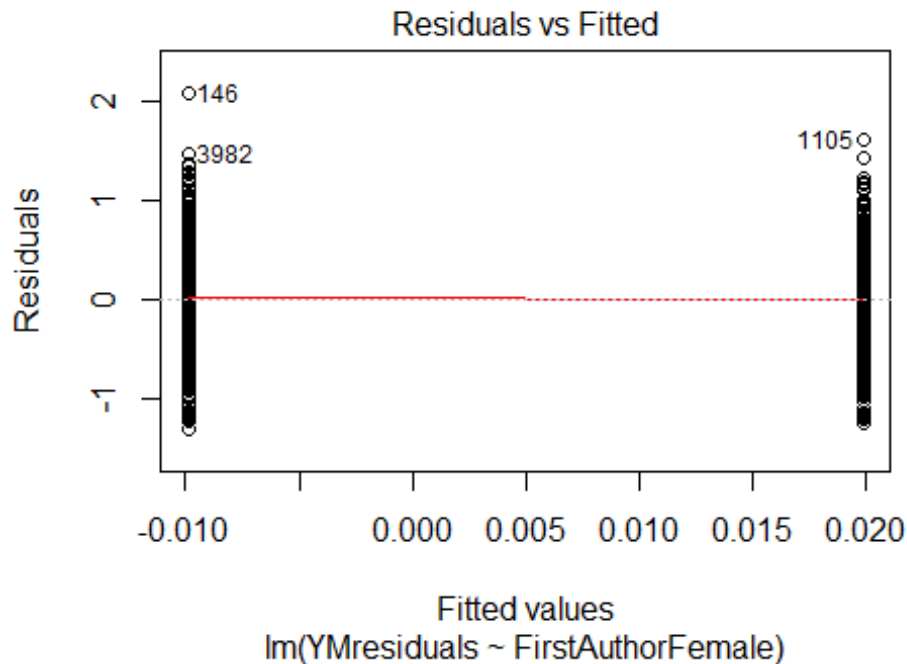
```



```
## 242 248
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 136 124 95 106 102 125 145 106 107 109 169 162 184 246 200
## 2011 2012
## 213 225
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 38, df = 16, p-value = 0.002
```

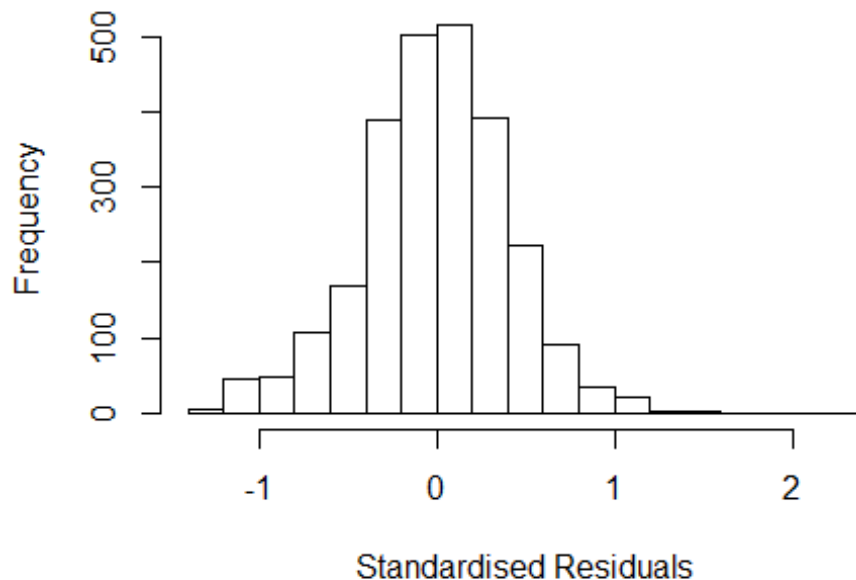


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 10, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 4.51106126563957"
## [1] "Male first author team size 2018 geometric mean: 4.36573095691014"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6700, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.23381825593883"
## [1] "Male last author team size 2018 geometric mean: 4.47485487053098"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4600, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.033 1          1.016
## LastAuthorFemale  1.091 1          1.044
## UniqueAuthors    1.175 4          1.020
## Year             1.237 16          1.007
```

## Residuals from first and last author and team size



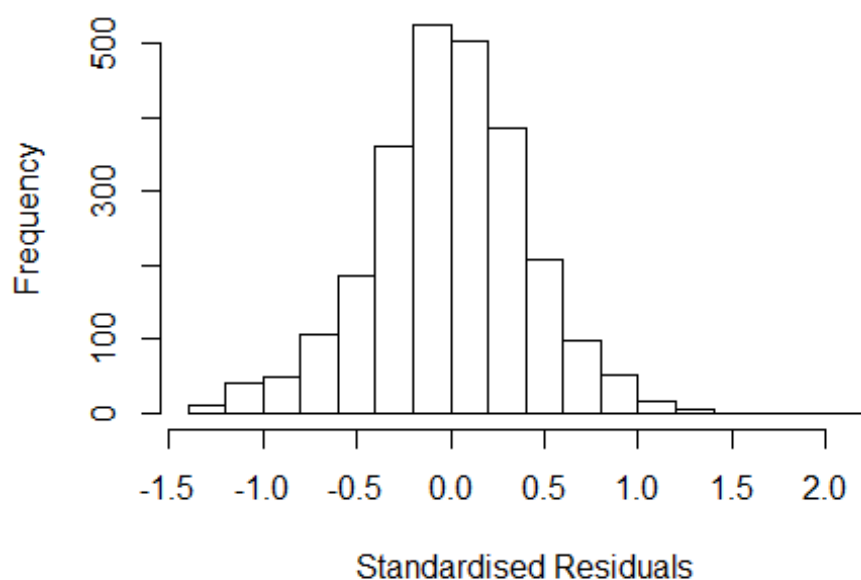
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29921 -0.24794  0.00262  0.25318  2.22835
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13965    0.05508   20.69 < 2e-16 ***
## FirstAuthorFemale1  0.03892    0.01670    2.33  0.0199 *
## LastAuthorFemale1 -0.03736    0.02011   -1.86  0.0633 .
## UniqueAuthors2     0.03285    0.04982    0.66  0.5097
## UniqueAuthors3     0.06581    0.04894    1.34  0.1788
## UniqueAuthors4     0.07455    0.04981    1.50  0.1346
## UniqueAuthors5     0.26130    0.04805    5.44 5.9e-08 ***
## Year1997         -0.00782    0.05558   -0.14  0.8881
## Year1998         -0.13887    0.05021   -2.77  0.0057 **
## Year1999         -0.04333    0.04905   -0.88  0.3771
```

```

## Year2000      -0.03743      0.05819      -0.64      0.5202
## Year2001      -0.09949      0.05125      -1.94      0.0524 .
## Year2002      -0.07520      0.04736      -1.59      0.1125
## Year2003      -0.11832      0.05440      -2.18      0.0297 *
## Year2004      -0.13668      0.05105      -2.68      0.0075 **
## Year2005      -0.09977      0.04762      -2.10      0.0363 *
## Year2006      -0.06532      0.04663      -1.40      0.1614
## Year2007      -0.12856      0.04748      -2.71      0.0068 **
## Year2008      -0.10174      0.04584      -2.22      0.0265 *
## Year2009      -0.10562      0.04275      -2.47      0.0136 *
## Year2010      -0.05359      0.04563      -1.17      0.2403
## Year2011      -0.19931      0.04752      -4.19      2.8e-05 ***
## Year2012      -0.13102      0.04793      -2.73      0.0063 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.381
## Multiple R-squared:  0.0795, Adjusted R-squared:  0.0715
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## observation 62 is an outlier with |weight| = 0 ( < 3.9e-05);
## 211 weights are ~= 1. The remaining 2342 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0921 0.8640 0.9520 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev mts compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.031 1 1.015
## LastAuthorFemale 1.073 1 1.036
## Year 1.081 16 1.002

```

## Residuals from first and last author



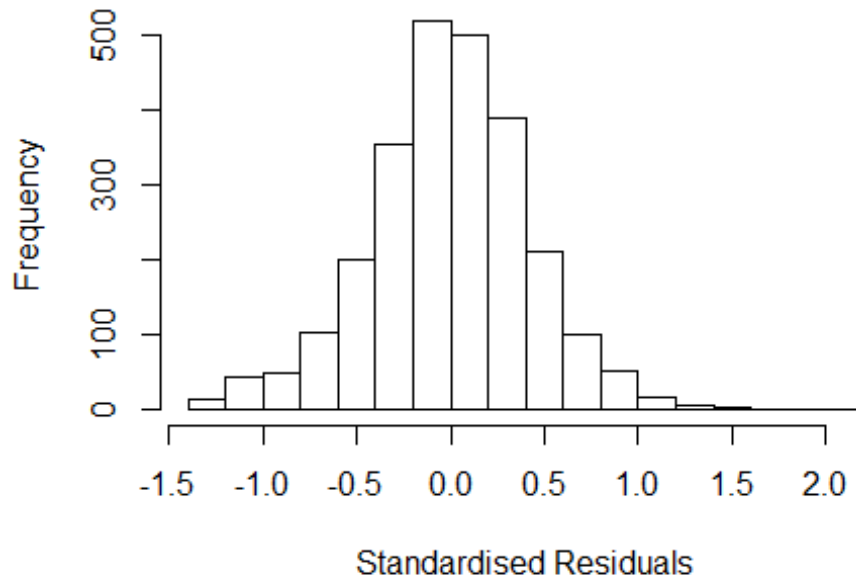
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25796 -0.25779 -0.00269 0.26386 2.09934
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2687 0.0375 33.85 < 2e-16 ***
## FirstAuthorFemale1 0.0364 0.0171 2.13 0.03343 *
## LastAuthorFemale1 -0.0578 0.0205 -2.82 0.00484 **
## Year1997 -0.0268 0.0569 -0.47 0.63829
## Year1998 -0.1402 0.0538 -2.61 0.00913 **
## Year1999 -0.0567 0.0500 -1.13 0.25691
## Year2000 -0.0261 0.0607 -0.43 0.66688
## Year2001 -0.0775 0.0530 -1.46 0.14362
## Year2002 -0.0619 0.0497 -1.24 0.21325
## Year2003 -0.0911 0.0533 -1.71 0.08734 .
## Year2004 -0.1099 0.0534 -2.06 0.03994 *
## Year2005 -0.0827 0.0490 -1.69 0.09174 .
```

```

## Year2006          -0.0508      0.0479   -1.06   0.28948
## Year2007          -0.1210      0.0489   -2.47   0.01341 *
## Year2008          -0.1019      0.0478   -2.13   0.03290 *
## Year2009          -0.0950      0.0443   -2.15   0.03199 *
## Year2010          -0.0471      0.0468   -1.01   0.31405
## Year2011          -0.1898      0.0497   -3.82   0.00014 ***
## Year2012          -0.1230      0.0499   -2.47   0.01374 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.0108
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 62 is an outlier with |weight| = 0 ( < 3.9e-05);
## 213 weights are ~ = 1. The remaining 2340 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0319 0.8660 0.9510 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.02 1          1.010
## Year              1.02 16          1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27e+00 -2.58e-01  8.47e-05  2.65e-01  2.10e+00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2650     0.0377   33.60 < 2e-16 ***
## FirstAuthorFemale1  0.0309     0.0170    1.81  0.0703 .
## Year1997        -0.0305     0.0569   -0.54  0.5920
## Year1998        -0.1469     0.0540   -2.72  0.0066 **
## Year1999        -0.0579     0.0501   -1.16  0.2479
## Year2000        -0.0246     0.0610   -0.40  0.6871
## Year2001        -0.0816     0.0532   -1.53  0.1254
## Year2002        -0.0614     0.0498   -1.23  0.2184
## Year2003        -0.0967     0.0535   -1.81  0.0709 .
## Year2004        -0.1114     0.0535   -2.08  0.0376 *
## Year2005        -0.0864     0.0493   -1.75  0.0797 .
## Year2006        -0.0608     0.0480   -1.27  0.2052
```

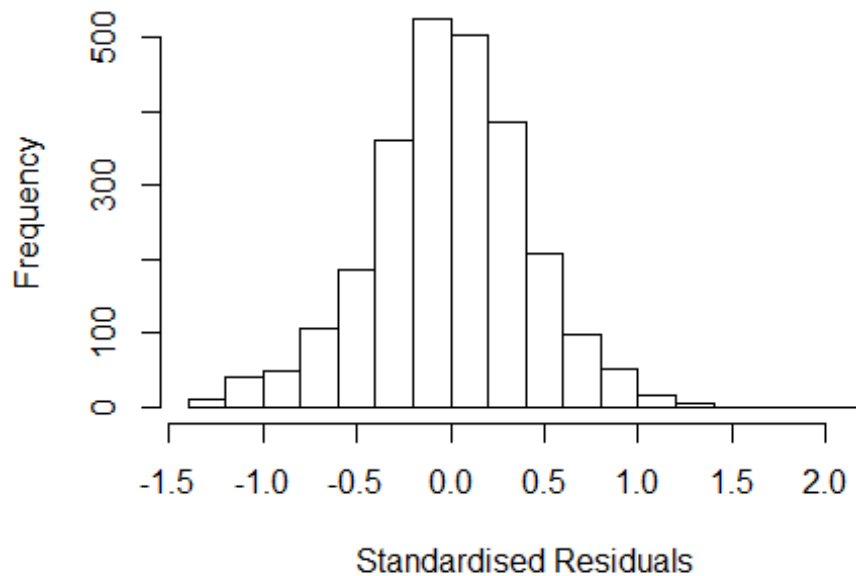
```

## Year2007          -0.1323      0.0487   -2.72   0.0067 **
## Year2008          -0.1074      0.0480   -2.24   0.0253 *
## Year2009          -0.1018      0.0444   -2.29   0.0218 *
## Year2010          -0.0503      0.0471   -1.07   0.2853
## Year2011          -0.1970      0.0500   -3.94   8.3e-05 ***
## Year2012          -0.1320      0.0501   -2.64   0.0085 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.392
## Multiple R-squared:  0.015, Adjusted R-squared:  0.0084
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 62 is an outlier with |weight| = 0 ( < 3.9e-05);
## 219 weights are ~= 1. The remaining 2334 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0534 0.8660 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.061 1          1.030
## Year            1.061 16          1.002

```



## Residuals from last author



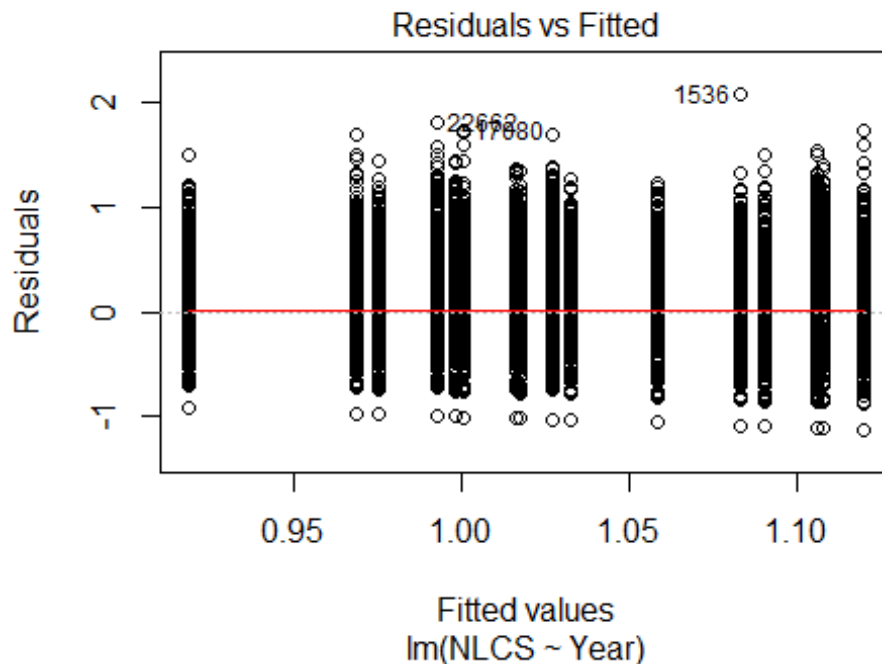
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.253786 -0.260101  0.000411  0.266044  2.088653
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2793     0.0370  34.62 < 2e-16 ***
## LastAuthorFemale1 -0.0524     0.0204  -2.56  0.01038 *
## Year1997         -0.0256     0.0569  -0.45  0.65325
## Year1998         -0.1418     0.0536  -2.64  0.00826 **
## Year1999         -0.0572     0.0500  -1.14  0.25262
## Year2000         -0.0284     0.0607  -0.47  0.63947
## Year2001         -0.0792     0.0529  -1.50  0.13411
## Year2002         -0.0642     0.0498  -1.29  0.19762
## Year2003         -0.0913     0.0534  -1.71  0.08727 .
## Year2004         -0.1100     0.0533  -2.06  0.03906 *
## Year2005         -0.0817     0.0491  -1.66  0.09647 .
## Year2006         -0.0497     0.0480  -1.04  0.30020
```

```

## Year2007          -0.1177      0.0491   -2.40   0.01653 *
## Year2008          -0.0982      0.0477   -2.06   0.03957 *
## Year2009          -0.0938      0.0442   -2.12   0.03402 *
## Year2010          -0.0471      0.0468   -1.01   0.31366
## Year2011          -0.1861      0.0495   -3.76   0.00017 ***
## Year2012          -0.1203      0.0498   -2.42   0.01573 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.391
## Multiple R-squared:  0.016, Adjusted R-squared:  0.00944
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 62 is an outlier with |weight| = 0 ( < 3.9e-05);
## 215 weights are ~= 1. The remaining 2338 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0259 0.8660 0.9510 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2554"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2738"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1085  990 1058 1017 1194 1236 1095 1059 1105 1199 1276 1270 1272 1339 1322
## 2011 2012
## 1483 1391
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  531  515  453  488  471  426  718  741  829  953 1000 1038  961 1050 1003

```

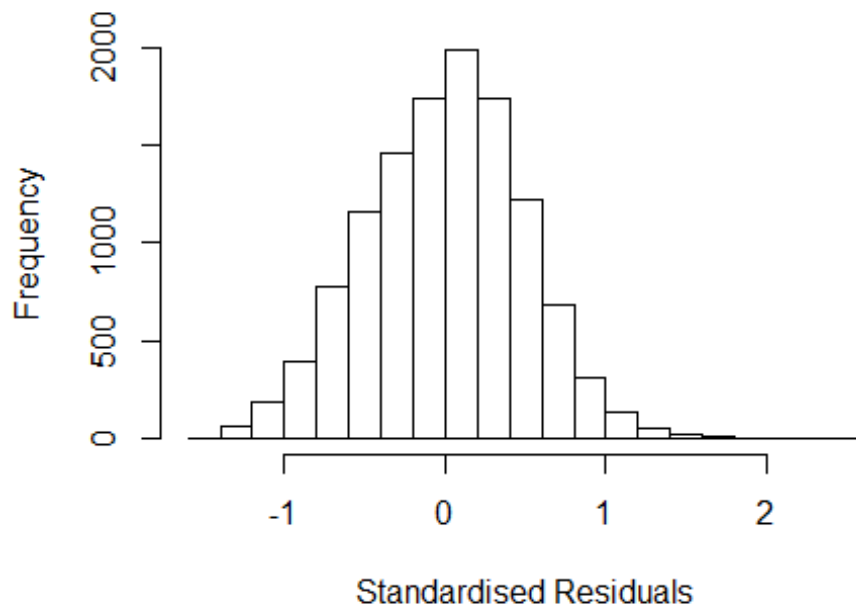
```
## 2011 2012
## 1163 1093
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 493 474 408 436 418 378 620 620 703 834 886 916 846 933 912
## 2011 2012
## 1052 1003
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 50, df = 16, p-value = 3e-05
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.1, df = 1, p-value = 0.7
```



## Residuals from first and last author and team size



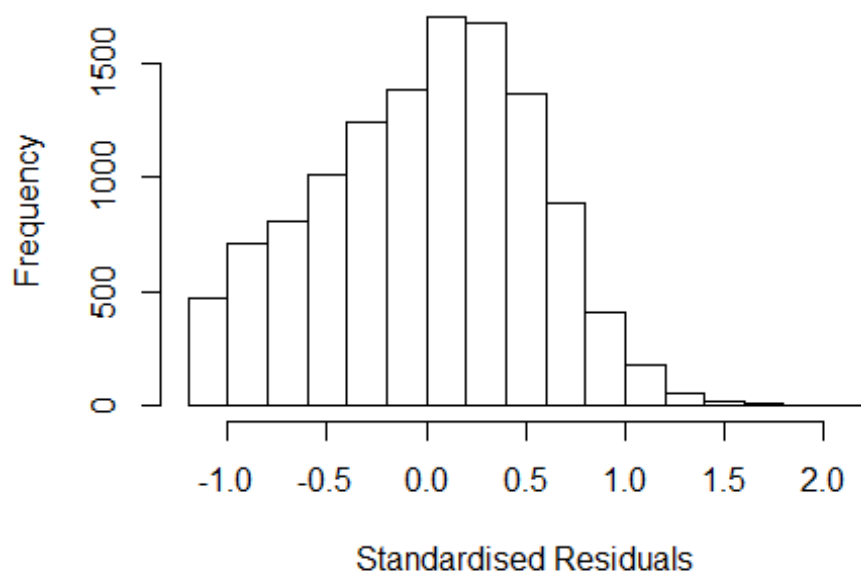
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4700 -0.3333 0.0199 0.3312 2.4281
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.80833 0.02647 30.53 < 2e-16 ***
## FirstAuthorFemale1 -0.00422 0.00968 -0.44 0.663
## LastAuthorFemale1 -0.04783 0.01003 -4.77 1.9e-06 ***
## UniqueAuthors2 0.34386 0.01440 23.89 < 2e-16 ***
## UniqueAuthors3 0.47888 0.01486 32.23 < 2e-16 ***
## UniqueAuthors4 0.59502 0.01585 37.55 < 2e-16 ***
## UniqueAuthors5 0.71350 0.01310 54.45 < 2e-16 ***
## Year1997 -0.07344 0.03513 -2.09 0.037 *
## Year1998 -0.02544 0.03744 -0.68 0.497
## Year1999 -0.05185 0.03487 -1.49 0.137
```

```

## Year2000      -0.07705      0.03465      -2.22      0.026 *
## Year2001      -0.08545      0.03581      -2.39      0.017 *
## Year2002      -0.12807      0.03252      -3.94      8.2e-05 ***
## Year2003      -0.18648      0.03171      -5.88      4.2e-09 ***
## Year2004      -0.20519      0.03056      -6.71      2.0e-11 ***
## Year2005      -0.26253      0.03042      -8.63      < 2e-16 ***
## Year2006      -0.22782      0.03018      -7.55      4.7e-14 ***
## Year2007      -0.19084      0.02944      -6.48      9.4e-11 ***
## Year2008      -0.20261      0.03069      -6.60      4.2e-11 ***
## Year2009      -0.20325      0.02998      -6.78      1.3e-11 ***
## Year2010      -0.19920      0.03033      -6.57      5.3e-11 ***
## Year2011      -0.21825      0.02974      -7.34      2.3e-13 ***
## Year2012      -0.19528      0.03037      -6.43      1.3e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.232, Adjusted R-squared:  0.23
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 625 is an outlier with |weight| = 0 ( < 8.4e-06);
## 946 weights are ~= 1. The remaining 10985 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0931 0.8710 0.9490 0.9080 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           8.38e-06           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.055 1 1.027
## LastAuthorFemale 1.043 1 1.021
## Year 1.023 16 1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1716 -0.4024 0.0424 0.3921 2.0846
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11817 0.02891 38.68 < 2e-16 ***
## FirstAuthorFemale1 0.05344 0.01085 4.92 8.6e-07 ***
## LastAuthorFemale1 -0.08589 0.01132 -7.59 3.5e-14 ***
## Year1997 -0.03975 0.03960 -1.00 0.31540
## Year1998 -0.00989 0.04283 -0.23 0.81743
## Year1999 -0.01671 0.03909 -0.43 0.66904
## Year2000 -0.01493 0.03901 -0.38 0.70187
## Year2001 -0.02778 0.03947 -0.70 0.48158
## Year2002 -0.06140 0.03542 -1.73 0.08308 .
## Year2003 -0.12494 0.03671 -3.40 0.00067 ***
## Year2004 -0.18197 0.03569 -5.10 3.5e-07 ***
## Year2005 -0.22859 0.03527 -6.48 9.5e-11 ***
```

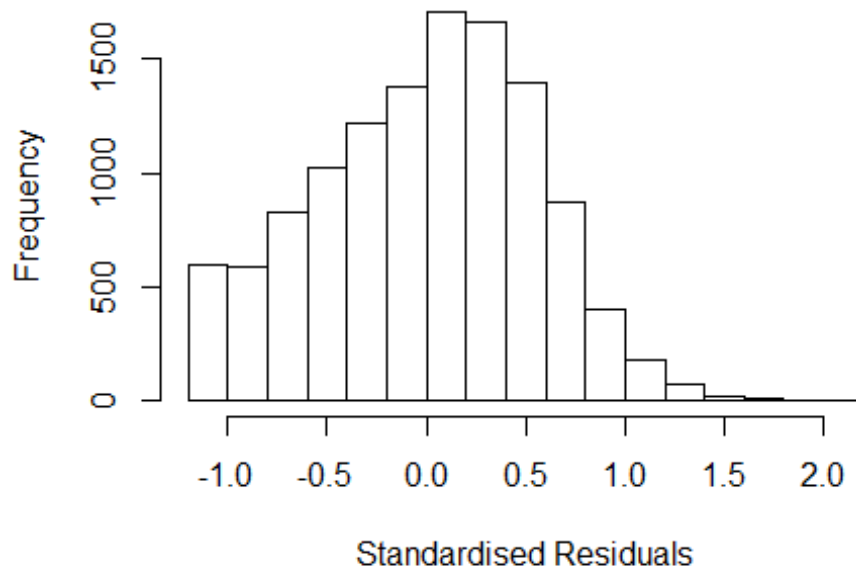
```

## Year2006          -0.16872    0.03455   -4.88  1.1e-06 ***
## Year2007          -0.08425    0.03364   -2.50  0.01229 *
## Year2008          -0.11915    0.03458   -3.45  0.00057 ***
## Year2009          -0.11735    0.03443   -3.41  0.00066 ***
## Year2010          -0.10691    0.03465   -3.09  0.00204 **
## Year2011          -0.12473    0.03441   -3.63  0.00029 ***
## Year2012          -0.08280    0.03439   -2.41  0.01607 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.0171, Adjusted R-squared:  0.0156
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 929 weights are ~= 1. The remaining 11003 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.152  0.875   0.948   0.914   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.38e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1      1.009
## Year              1.018 16      1.001

```



## Residuals from first author



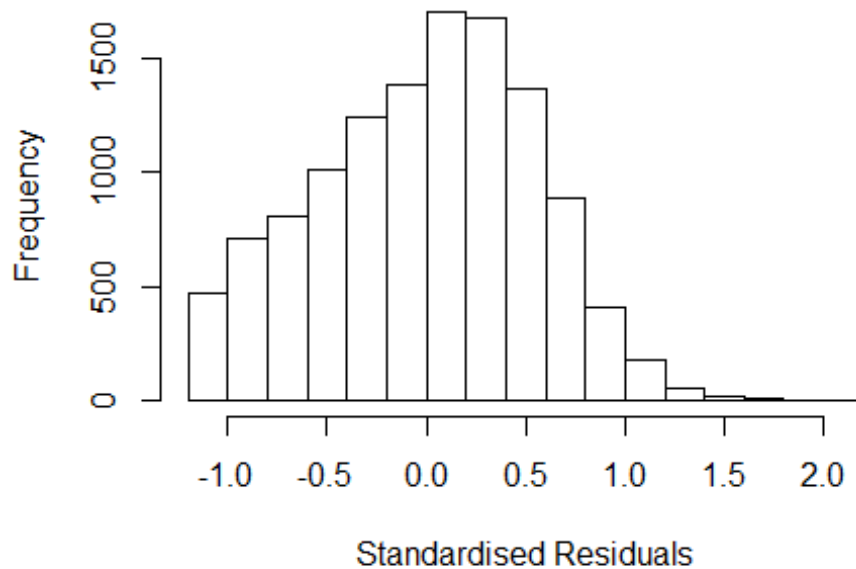
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1291 -0.4051 0.0437 0.3943 2.1063
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09793 0.02875 38.19 < 2e-16 ***
## FirstAuthorFemale1 0.03115 0.01076 2.90 0.00379 **
## Year1997 -0.04119 0.03958 -1.04 0.29807
## Year1998 -0.00659 0.04293 -0.15 0.87795
## Year1999 -0.01525 0.03907 -0.39 0.69623
## Year2000 -0.01269 0.03898 -0.33 0.74487
## Year2001 -0.02743 0.03943 -0.70 0.48668
## Year2002 -0.06133 0.03541 -1.73 0.08331 .
## Year2003 -0.12639 0.03663 -3.45 0.00056 ***
## Year2004 -0.18204 0.03576 -5.09 3.6e-07 ***
## Year2005 -0.23098 0.03532 -6.54 6.4e-11 ***
## Year2006 -0.16919 0.03454 -4.90 9.8e-07 ***
```

```

## Year2007          -0.08585    0.03365   -2.55  0.01074 *
## Year2008          -0.12059    0.03457   -3.49  0.00049 ***
## Year2009          -0.12148    0.03451   -3.52  0.00043 ***
## Year2010          -0.10865    0.03468   -3.13  0.00174 **
## Year2011          -0.12600    0.03446   -3.66  0.00026 ***
## Year2012          -0.08698    0.03440   -2.53  0.01146 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.0122, Adjusted R-squared:  0.0108
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 936 weights are ~= 1. The remaining 10996 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.143  0.873  0.948  0.914  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.38e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.006 1          1.003
## Year            1.006 16          1.000

```

## Residuals from last author



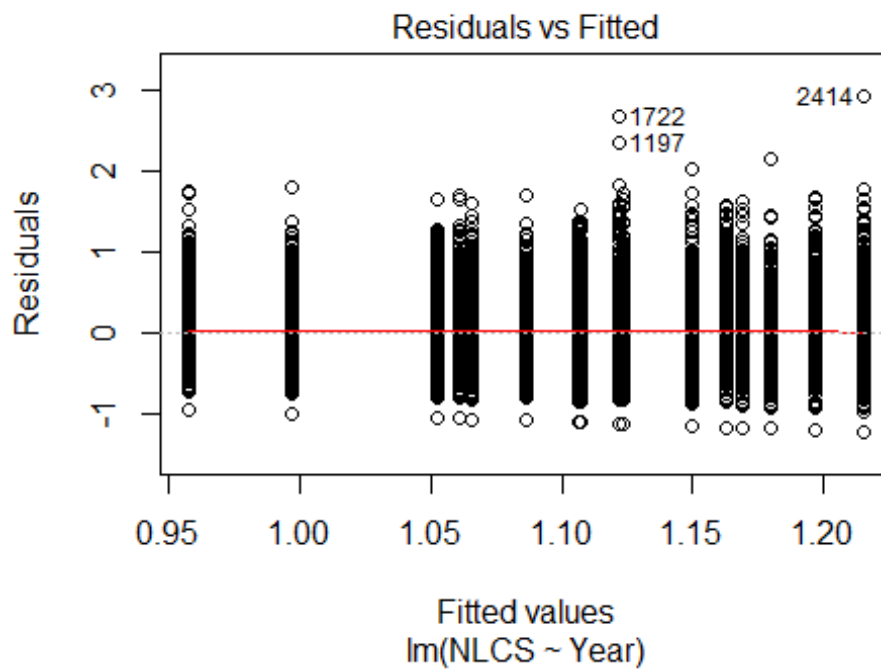
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1318 -0.4045 0.0421 0.3937 2.0702
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13180 0.02870 39.44 < 2e-16 ***
## LastAuthorFemale1 -0.07113 0.01111 -6.41 1.6e-10 ***
## Year1997 -0.03901 0.03957 -0.99 0.32423
## Year1998 -0.00776 0.04281 -0.18 0.85621
## Year1999 -0.01214 0.03903 -0.31 0.75585
## Year2000 -0.01045 0.03896 -0.27 0.78858
## Year2001 -0.02170 0.03950 -0.55 0.58267
## Year2002 -0.05698 0.03536 -1.61 0.10707
## Year2003 -0.11934 0.03664 -3.26 0.00113 **
## Year2004 -0.17614 0.03565 -4.94 7.9e-07 ***
## Year2005 -0.22470 0.03527 -6.37 1.9e-10 ***
## Year2006 -0.16275 0.03446 -4.72 2.4e-06 ***
```

```

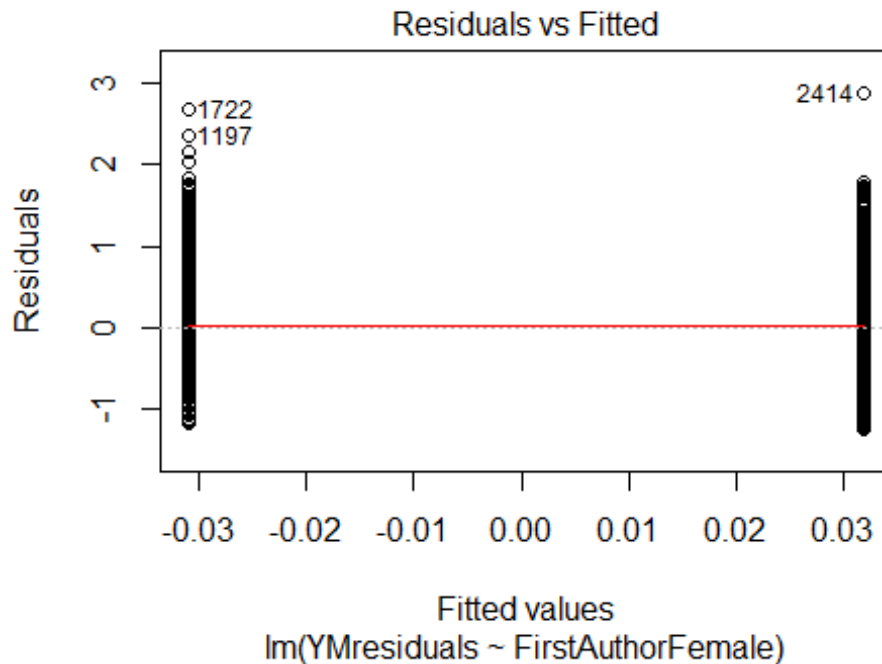
## Year2007          -0.07818      0.03358    -2.33  0.01992 *
## Year2008          -0.11157      0.03448    -3.24  0.00122 **
## Year2009          -0.11067      0.03436    -3.22  0.00128 **
## Year2010          -0.09950      0.03460    -2.88  0.00404 **
## Year2011          -0.11433      0.03428    -3.33  0.00086 ***
## Year2012          -0.07392      0.03431    -2.15  0.03124 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.0151, Adjusted R-squared:  0.0137
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 949 weights are ~ = 1. The remaining 10983 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.161  0.876  0.948  0.915  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.38e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 11932"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2739"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  920 1000  996  922 1119 1074 1030  969  990 1133 1229 1314 1309 1265 1296
## 2011 2012
## 1363 1400
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  424 465 447 433 479 392 593 619 642 829 865 973 936 857 849
## 2011 2012

```

```
## 928 916
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 397 430 414 390 440 362 506 522 561 729 761 840 801 787 780
## 2011 2012
## 835 832
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 72, df = 16, p-value = 4e-09
```

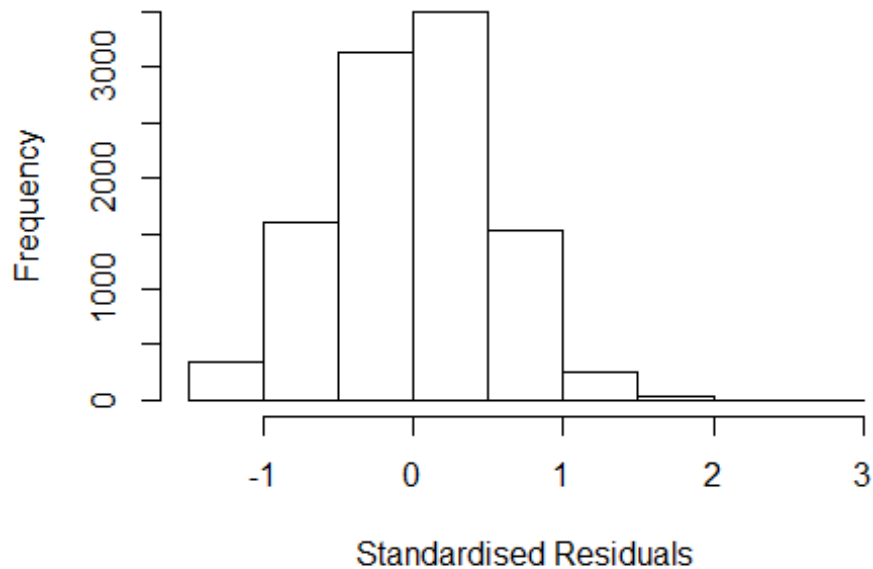


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 3e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.5513814222658"
## [1] "Male first author team size 2018 geometric mean: 3.32254936986246"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.46126474030134"
## [1] "Male last author team size 2018 geometric mean: 3.44976793137215"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1          1.041
## LastAuthorFemale  1.076 1          1.038
## UniqueAuthors     1.079 4          1.010
## Year              1.069 16          1.002
```

## Residuals from first and last author and team size



```
## [1] "List of 3 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1197 0031466983 3.453 1997    1100     5    2.567
## 1722 0031279593 3.788 1997    2739     1    2.902
## 2414 0031664637 4.128 1998    2739     3    2.828
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4108 -0.3620  0.0159  0.3693  2.9016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96021    0.03490   27.51 < 2e-16 ***
## FirstAuthorFemale1 0.03419    0.01139    3.00 0.00269 **
## LastAuthorFemale1 0.01047    0.01168    0.90 0.37033
## UniqueAuthors2    0.23812    0.01776   13.41 < 2e-16 ***
## UniqueAuthors3    0.32232    0.01738   18.54 < 2e-16 ***
## UniqueAuthors4    0.38936    0.01936   20.12 < 2e-16 ***
## UniqueAuthors5    0.43315    0.01659   26.10 < 2e-16 ***
## Year1997        -0.07378    0.04530   -1.63 0.10343
```

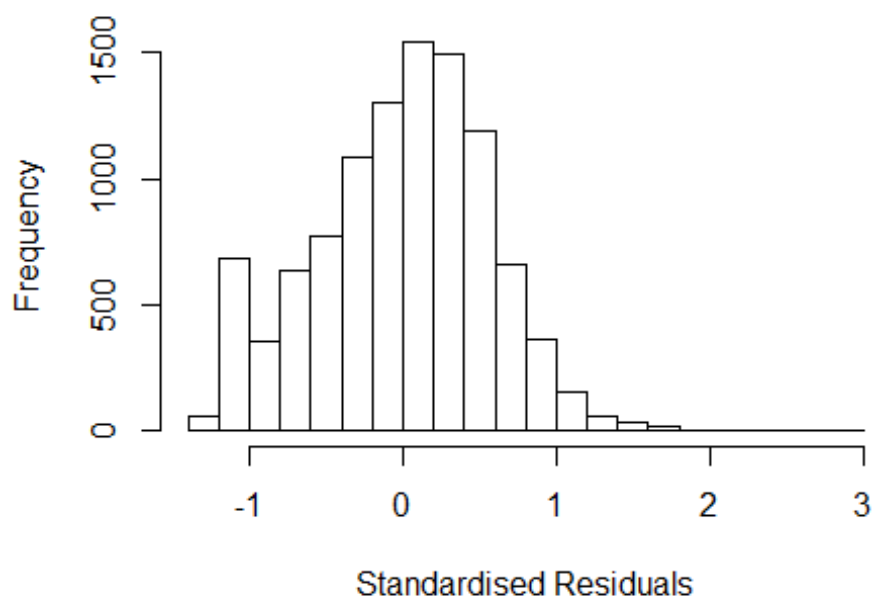
```

## Year1998      -0.01678    0.04395   -0.38  0.70266
## Year1999      -0.00906    0.04316   -0.21  0.83372
## Year2000      -0.02029    0.04300   -0.47  0.63709
## Year2001      -0.03824    0.04574   -0.84  0.40317
## Year2002      -0.10644    0.04162   -2.56  0.01056 *
## Year2003      -0.17241    0.04165   -4.14  3.5e-05 ***
## Year2004      -0.16701    0.04108   -4.07  4.8e-05 ***
## Year2005      -0.27134    0.03962   -6.85  7.9e-12 ***
## Year2006      -0.22678    0.03926   -5.78  7.8e-09 ***
## Year2007      -0.18731    0.03805   -4.92  8.7e-07 ***
## Year2008      -0.18069    0.03816   -4.74  2.2e-06 ***
## Year2009      -0.12606    0.03858   -3.27  0.00109 **
## Year2010      -0.14235    0.03792   -3.75  0.00017 ***
## Year2011      -0.13311    0.03782   -3.52  0.00043 ***
## Year2012      -0.09676    0.03831   -2.53  0.01157 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0956, Adjusted R-squared:  0.0937
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 3 observations c(469,644,908) are outliers with |weight| = 0 ( < 9.6e-
06);
## 872 weights are ~= 1. The remaining 9512 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.117  0.866  0.950  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.63e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.057 1          1.028
## LastAuthorFemale 1.057 1          1.028
## Year 1.010 16          1.000

```



## Residuals from first and last author



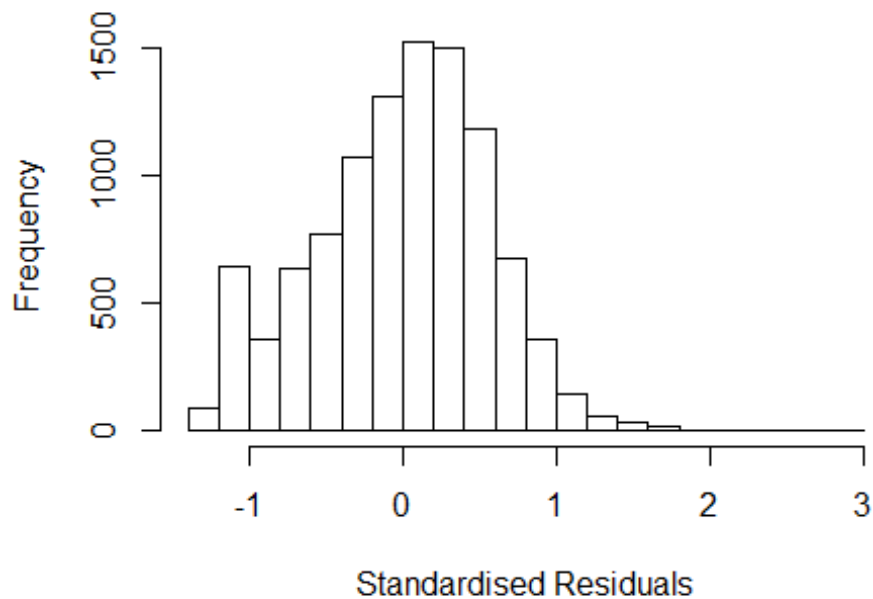
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1722 0031279593 3.788 1997      2739      1      2.700
## 2414 0031664637 4.128 1998      2739      3      2.888
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2400 -0.3777  0.0409  0.3820  2.8880
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.138991   0.035264  32.30 < 2e-16 ***
## FirstAuthorFemale1 0.076092   0.011767   6.47 1.0e-10 ***
## LastAuthorFemale1 -0.019327   0.012096  -1.60  0.1101
## Year1997        -0.051225   0.047643  -1.08  0.2823
## Year1998         0.024875   0.045668   0.54  0.5860
## Year1999         0.028289   0.045307   0.62  0.5324
## Year2000         0.035053   0.044511   0.79  0.4310
## Year2001         0.008289   0.046679   0.18  0.8591
## Year2002        -0.049237   0.043260  -1.14  0.2551
## Year2003        -0.109376   0.043371  -2.52  0.0117 *
## Year2004        -0.104109   0.042724  -2.44  0.0148 *
```

```

## Year2005          -0.213680    0.042048    -5.08    3.8e-07 ***
## Year2006          -0.164077    0.041378    -3.97    7.4e-05 ***
## Year2007          -0.106706    0.039855    -2.68    0.0074 **
## Year2008          -0.112939    0.039875    -2.83    0.0046 **
## Year2009          -0.045222    0.040068    -1.13    0.2591
## Year2010          -0.059542    0.039650    -1.50    0.1332
## Year2011          -0.036034    0.039847    -0.90    0.3658
## Year2012           0.000299    0.040114    0.01    0.9940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.018, Adjusted R-squared:  0.0163
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(644,908) are outliers with |weight| = 0 ( < 9.6e-06);
## 843 weights are ~ = 1. The remaining 9542 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0254 0.8630 0.9480 0.9040 0.9840 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          9.63e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.006 1           1.003
## Year              1.006 16           1.000

```

## Residuals from first author



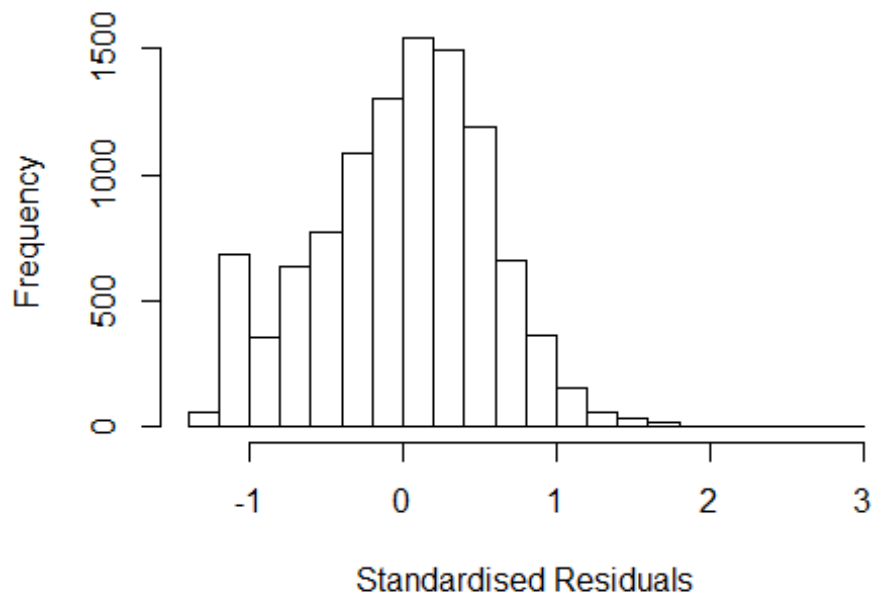
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1722 0031279593 3.788 1997      2739      1      2.700
## 2414 0031664637 4.128 1998      2739      3      2.888
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2397 -0.3763  0.0414  0.3835  2.8984
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.134919   0.035131  32.31  < 2e-16 ***
## FirstAuthorFemale1 0.069847   0.011555   6.04  1.5e-09 ***
## Year1997       -0.050940   0.047670  -1.07   0.2853
## Year1998        0.024800   0.045638   0.54   0.5869
## Year1999        0.027504   0.045310   0.61   0.5439
## Year2000        0.034940   0.044496   0.79   0.4323
## Year2001        0.008134   0.046661   0.17   0.8616
## Year2002       -0.049502   0.043243  -1.14   0.2523
## Year2003       -0.109270   0.043374  -2.52   0.0118 *
## Year2004       -0.105070   0.042752  -2.46   0.0140 *
## Year2005       -0.213674   0.042048  -5.08  3.8e-07 ***
```

```

## Year2006          -0.164374    0.041367    -3.97    7.1e-05 ***
## Year2007          -0.106413    0.039849    -2.67    0.0076 **
## Year2008          -0.113343    0.039864    -2.84    0.0045 **
## Year2009          -0.045215    0.040056    -1.13    0.2590
## Year2010          -0.060158    0.039649    -1.52    0.1292
## Year2011          -0.037162    0.039816    -0.93    0.3507
## Year2012          -0.000431    0.040093    -0.01    0.9914
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0178, Adjusted R-squared:  0.0162
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(644,908) are outliers with |weight| = 0 ( < 9.6e-06);
## 843 weights are ~ = 1. The remaining 9542 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0245 0.8630 0.9490 0.9040 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.63e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.006 1          1.003
## Year            1.006 16          1.000

```

## Residuals from last author



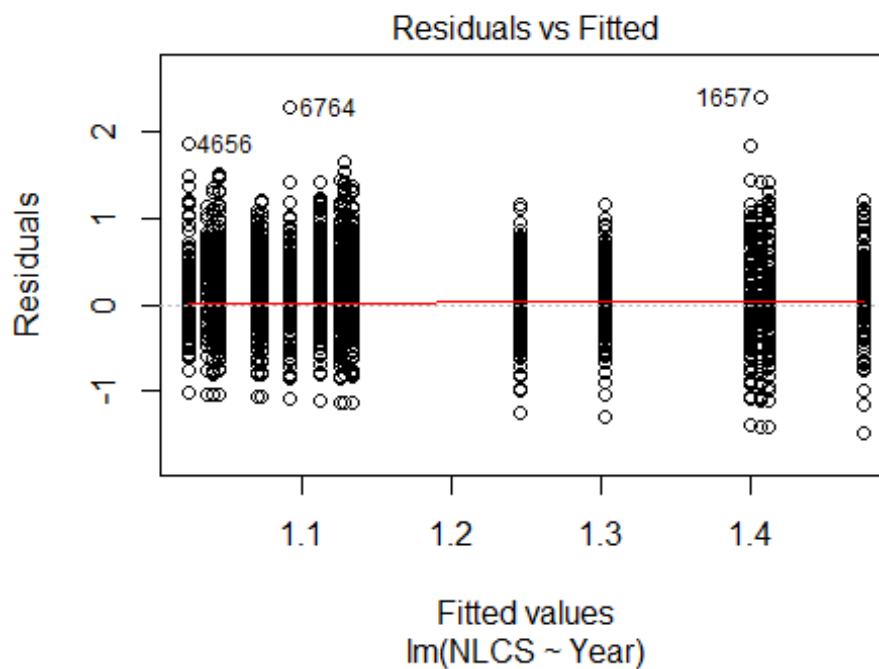
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1722 0031279593 3.788 1997      2739      1      2.700
## 2414 0031664637 4.128 1998      2739      3      2.888
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2088 -0.3815  0.0433  0.3828  2.9379
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16267    0.03512   33.11 < 2e-16 ***
## LastAuthorFemale1 0.00643    0.01187    0.54  0.58772
## Year1997      -0.05339    0.04782   -1.12  0.26430
## Year1998       0.02738    0.04569    0.60  0.54897
## Year1999       0.03096    0.04536    0.68  0.49484
## Year2000       0.03974    0.04466    0.89  0.37353
## Year2001       0.01402    0.04679    0.30  0.76440
## Year2002      -0.04610    0.04345   -1.06  0.28872
## Year2003      -0.10438    0.04350   -2.40  0.01644 *
## Year2004      -0.10228    0.04296   -2.38  0.01729 *
## Year2005      -0.20702    0.04213   -4.91  9.1e-07 ***
```

```

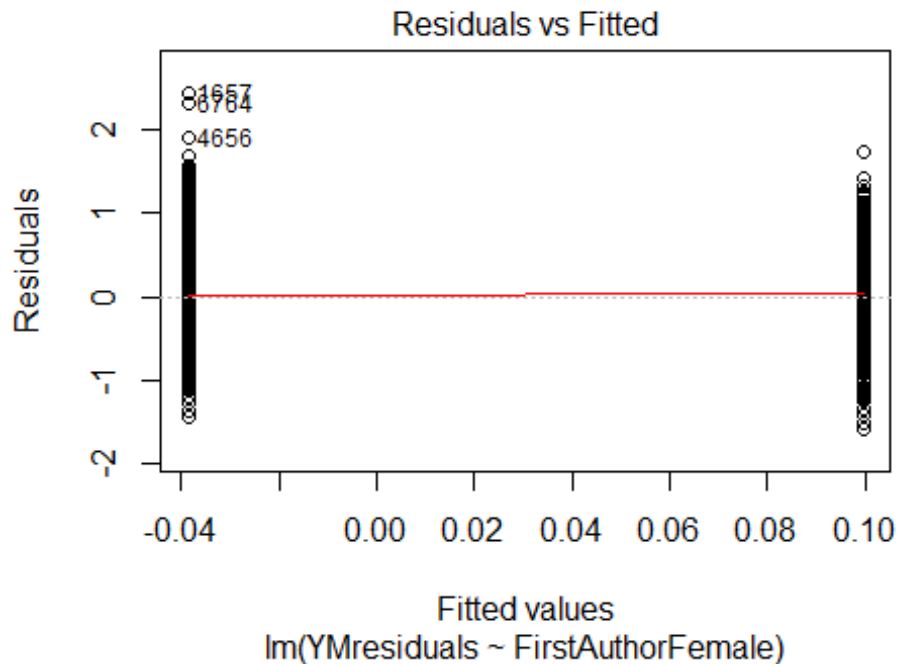
## Year2006          -0.16014      0.04143    -3.87  0.00011 ***
## Year2007          -0.10125      0.03998    -2.53  0.01134 *
## Year2008          -0.10855      0.04000    -2.71  0.00666 **
## Year2009          -0.03941      0.04016    -0.98  0.32654
## Year2010          -0.05162      0.03977    -1.30  0.19434
## Year2011          -0.03024      0.03996    -0.76  0.44931
## Year2012           0.00697      0.04021     0.17  0.86247
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0141, Adjusted R-squared:  0.0125
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 2 observations c(644,908) are outliers with |weight| = 0 ( < 9.6e-06);
## 858 weights are ~ = 1. The remaining 9527 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0313 0.8620 0.9480 0.9040 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.63e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 10387"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2740"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 597 601 635 608 717 314 422 328 323 367 344 438 387 463 423
## 2011 2012
## 446 464
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 120 117 125 136 165 95 200 157 179 209 221 321 221 261 287
## 2011 2012
## 303 346
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 107 105 115 122 151 85 159 129 142 175 192 277 196 225 254
## 2011 2012
## 259 297
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 57, df = 16, p-value = 2e-06
```



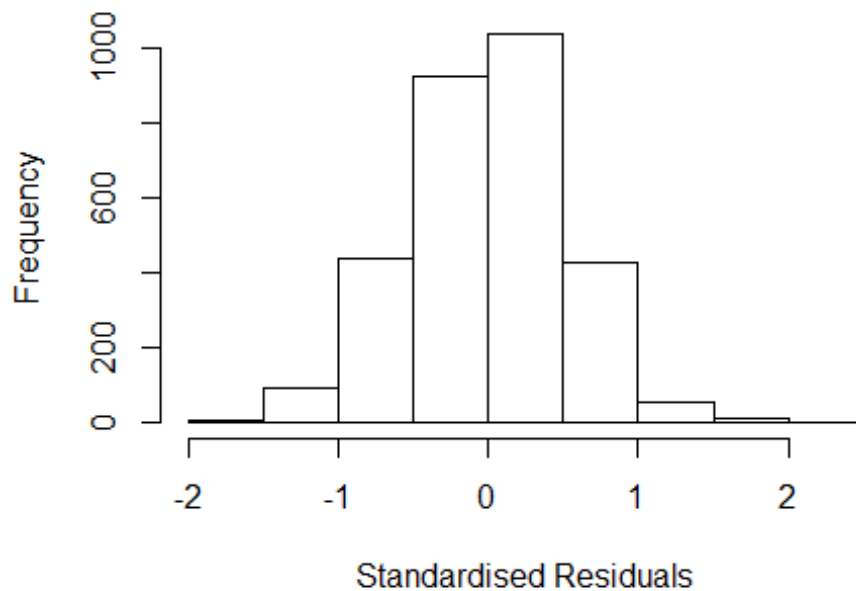
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.67, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 5.35936649993976"
## [1] "Male first author team size 2018 geometric mean: 4.89899712596899"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.17060557661002"
## [1] "Male last author team size 2018 geometric mean: 5.06714796817764"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7100, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.083 1          1.041
## LastAuthorFemale  1.049 1          1.024
## UniqueAuthors    1.105 4          1.013
## Year             1.159 16         1.005
```



## Residuals from first and last author and team size



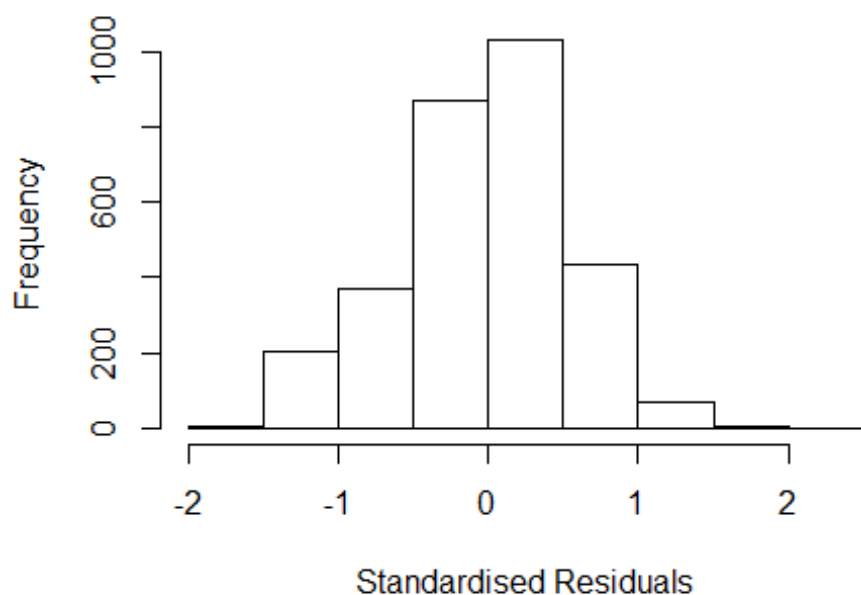
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7777 -0.3533 0.0112 0.3516 2.0338
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9943 0.0714 13.93 < 2e-16 ***
## FirstAuthorFemale1 0.0993 0.0221 4.50 7.0e-06 ***
## LastAuthorFemale1 0.1027 0.0287 3.58 0.00035 ***
## UniqueAuthors2 0.2644 0.0559 4.73 2.3e-06 ***
## UniqueAuthors3 0.3013 0.0524 5.75 9.9e-09 ***
## UniqueAuthors4 0.3670 0.0510 7.20 7.5e-13 ***
## UniqueAuthors5 0.6432 0.0482 13.34 < 2e-16 ***
## Year1997 0.0410 0.0795 0.52 0.60621
## Year1998 0.0260 0.0839 0.31 0.75653
## Year1999 -0.0113 0.0791 -0.14 0.88650
```

```

## Year2000      -0.1233      0.0708      -1.74      0.08185 .
## Year2001      -0.4670      0.0843      -5.54      3.3e-08 ***
## Year2002      -0.2515      0.0683      -3.68      0.00024 ***
## Year2003      -0.4544      0.0734      -6.19      6.8e-10 ***
## Year2004      -0.3748      0.0714      -5.25      1.6e-07 ***
## Year2005      -0.4370      0.0680      -6.43      1.5e-10 ***
## Year2006      -0.3702      0.0679      -5.45      5.4e-08 ***
## Year2007      -0.3528      0.0646      -5.46      5.0e-08 ***
## Year2008      -0.3965      0.0667      -5.95      3.0e-09 ***
## Year2009      -0.4350      0.0685      -6.35      2.5e-10 ***
## Year2010      -0.3538      0.0682      -5.19      2.3e-07 ***
## Year2011      -0.3900      0.0686      -5.68      1.5e-08 ***
## Year2012      -0.3702      0.0661      -5.60      2.3e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.19, Adjusted R-squared:  0.184
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 251 weights are ~= 1. The remaining 2739 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0952 0.8730 0.9500 0.9060 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1      1.028
## LastAuthorFemale 1.040 1      1.020
## Year      1.072 16      1.002

```

## Residuals from first and last author



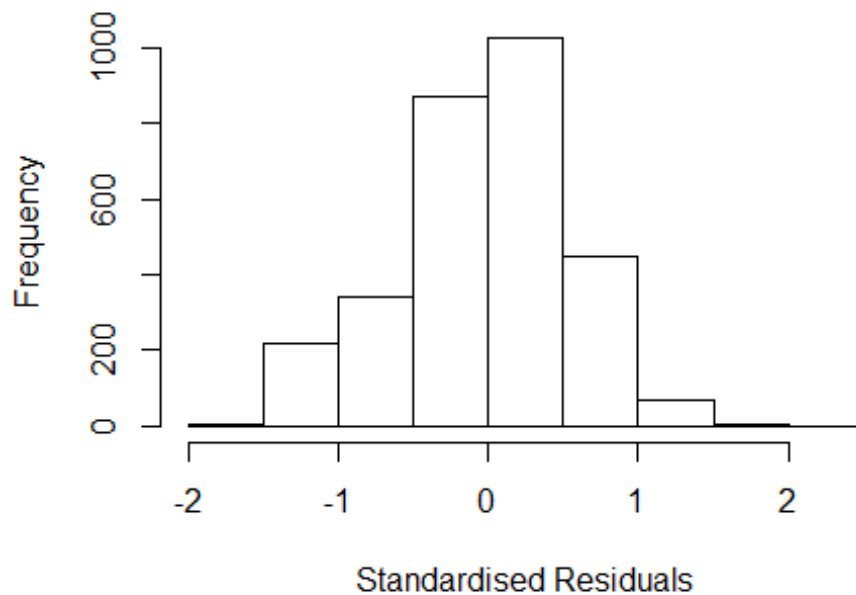
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5949 -0.3594 0.0286 0.3640 2.3182
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.39459 0.06117 22.80 < 2e-16 ***
## FirstAuthorFemale1 0.12898 0.02349 5.49 4.3e-08 ***
## LastAuthorFemale1 0.10267 0.03155 3.25 0.0012 **
## Year1997 0.07137 0.08088 0.88 0.3776
## Year1998 -0.01542 0.08702 -0.18 0.8593
## Year1999 0.00553 0.08411 0.07 0.9476
## Year2000 -0.12244 0.07541 -1.62 0.1046
## Year2001 -0.41267 0.08927 -4.62 3.9e-06 ***
## Year2002 -0.19557 0.07305 -2.68 0.0075 **
## Year2003 -0.42213 0.08030 -5.26 1.6e-07 ***
## Year2004 -0.36241 0.07480 -4.85 1.3e-06 ***
## Year2005 -0.41308 0.07219 -5.72 1.2e-08 ***
```

```

## Year2006      -0.34968    0.07035   -4.97  7.1e-07 ***
## Year2007      -0.30399    0.06875   -4.42  1.0e-05 ***
## Year2008      -0.35774    0.07086   -5.05  4.7e-07 ***
## Year2009      -0.40163    0.07258   -5.53  3.4e-08 ***
## Year2010      -0.30120    0.07432   -4.05  5.2e-05 ***
## Year2011      -0.34460    0.07403   -4.65  3.4e-06 ***
## Year2012      -0.31466    0.07143   -4.41  1.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0739, Adjusted R-squared:  0.0683
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 241 weights are ~= 1. The remaining 2749 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0276 0.8640 0.9510 0.9010 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1      1.020
## Year      1.041 16      1.001

```

## Residuals from first author



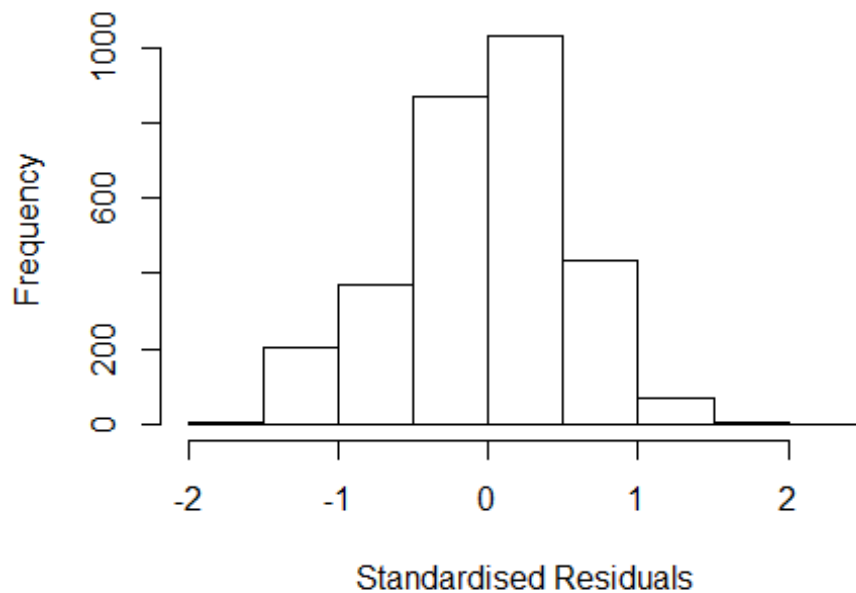
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.614 -0.365 0.033 0.368 2.415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.40577 0.06068 23.17 < 2e-16 ***
## FirstAuthorFemale1 0.14171 0.02335 6.07 1.4e-09 ***
## Year1997 0.06607 0.08079 0.82 0.4135
## Year1998 -0.02071 0.08667 -0.24 0.8112
## Year1999 0.00746 0.08358 0.09 0.9289
## Year2000 -0.12549 0.07481 -1.68 0.0936 .
## Year2001 -0.41520 0.08921 -4.65 3.4e-06 ***
## Year2002 -0.19453 0.07280 -2.67 0.0076 **
## Year2003 -0.42448 0.08002 -5.30 1.2e-07 ***
## Year2004 -0.36649 0.07472 -4.90 9.9e-07 ***
## Year2005 -0.41348 0.07208 -5.74 1.1e-08 ***
## Year2006 -0.35114 0.07021 -5.00 6.0e-07 ***
```

```

## Year2007          -0.30176      0.06844      -4.41  1.1e-05 ***
## Year2008          -0.35928      0.07051      -5.10  3.7e-07 ***
## Year2009          -0.40252      0.07227      -5.57  2.8e-08 ***
## Year2010          -0.29876      0.07442      -4.01  6.1e-05 ***
## Year2011          -0.34162      0.07378      -4.63  3.8e-06 ***
## Year2012          -0.31282      0.07102      -4.40  1.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0699, Adjusted R-squared:  0.0645
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 242 weights are ~= 1. The remaining 2748 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0102 0.8640 0.9510 0.9010 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.629 -0.369 0.034 0.373 2.257
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.42052 0.06074 23.39 < 2e-16 ***
## LastAuthorFemale1 0.12893 0.03120 4.13 3.7e-05 ***
## Year1997 0.07947 0.08064 0.99 0.3245
## Year1998 -0.00659 0.08633 -0.08 0.9391
## Year1999 0.00766 0.08397 0.09 0.9273
## Year2000 -0.12202 0.07473 -1.63 0.1026
## Year2001 -0.40222 0.08892 -4.52 6.3e-06 ***
## Year2002 -0.19117 0.07285 -2.62 0.0087 **
## Year2003 -0.42007 0.08020 -5.24 1.7e-07 ***
## Year2004 -0.35524 0.07446 -4.77 1.9e-06 ***
## Year2005 -0.41784 0.07205 -5.80 7.4e-09 ***
## Year2006 -0.35119 0.07037 -4.99 6.4e-07 ***
```

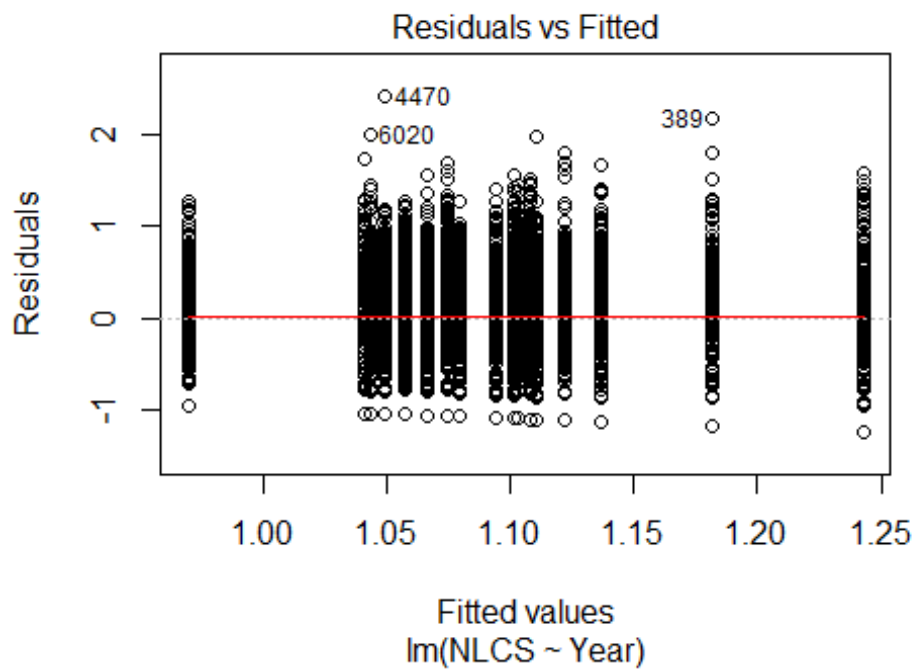
```

## Year2007          -0.30100      0.06862    -4.39  1.2e-05 ***
## Year2008          -0.33628      0.07044    -4.77  1.9e-06 ***
## Year2009          -0.38949      0.07229    -5.39  7.7e-08 ***
## Year2010          -0.29517      0.07437    -3.97  7.4e-05 ***
## Year2011          -0.32951      0.07373    -4.47  8.2e-06 ***
## Year2012          -0.29880      0.07106    -4.20  2.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0634, Adjusted R-squared:  0.058
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 236 weights are ~= 1. The remaining 2754 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0472 0.8670 0.9490 0.9010 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.34e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2990"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2741"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 675 710 744 824 858 892 739 638 692 722 770 830 830 848 761
## 2011 2012
## 766 787
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 178 209 205 239 234 235 286 265 326 358 374 428 418 475 435
## 2011 2012

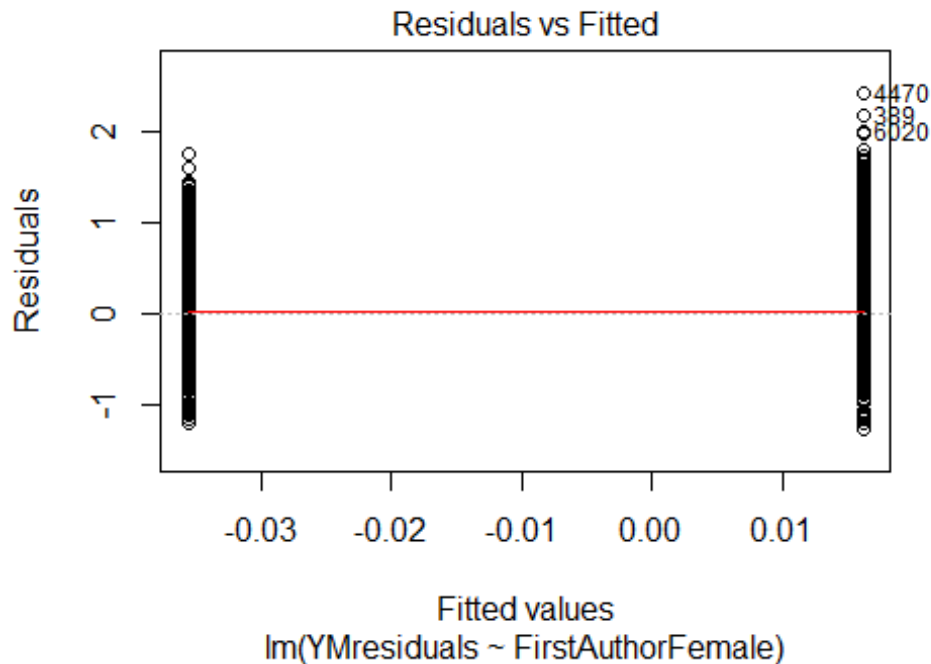
```



```
## 442 445
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 160 180 184 215 211 208 257 187 249 264 295 325 339 384 375
## 2011 2012
## 380 319
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 67, df = 16, p-value = 4e-08
```

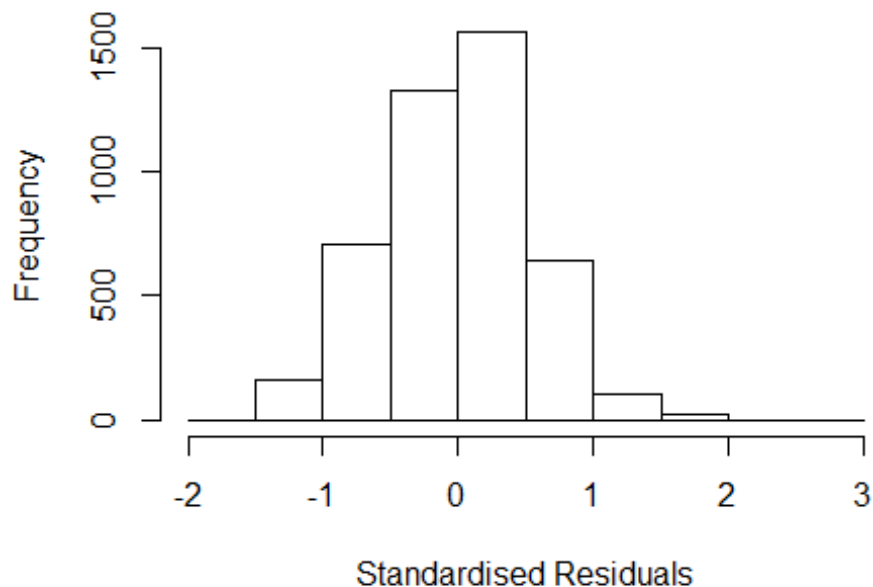


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 5.10927747878775"
## [1] "Male first author team size 2018 geometric mean: 5.47258202656695"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.04713756938061"
## [1] "Male last author team size 2018 geometric mean: 5.43174356818876"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.073 1 1.036
## LastAuthorFemale 1.094 1 1.046
## UniqueAuthors 1.135 4 1.016
## Year 1.120 16 1.004
```

## Residuals from first and last author and team size



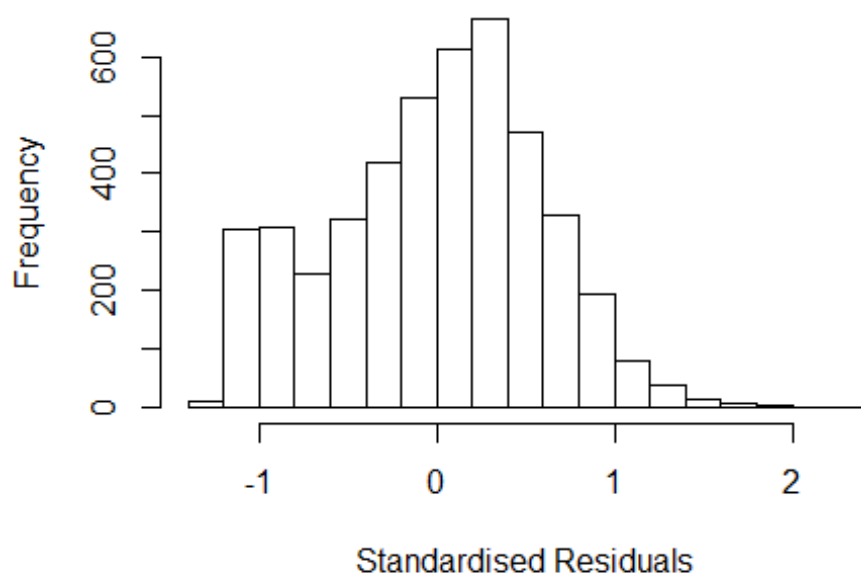
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4470 0035384136 3.474 2001      1704      5      2.587
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5696 -0.3852  0.0209  0.3696  2.5867
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7645    0.0573   13.33 < 2e-16 ***
## FirstAuthorFemale1 -0.0348    0.0181   -1.93  0.05423 .
## LastAuthorFemale1 -0.1001    0.0216   -4.63  3.7e-06 ***
## UniqueAuthors2     0.3402    0.0371    9.16 < 2e-16 ***
## UniqueAuthors3     0.4606    0.0356   12.94 < 2e-16 ***
## UniqueAuthors4     0.5229    0.0360   14.54 < 2e-16 ***
## UniqueAuthors5     0.7931    0.0314   25.23 < 2e-16 ***
## Year1997          0.0120    0.0722    0.17  0.86754
## Year1998         -0.1083    0.0682   -1.59  0.11262
## Year1999         -0.0600    0.0660   -0.91  0.36373
```

```

## Year2000          -0.1493      0.0642    -2.33  0.02001 *
## Year2001          -0.2174      0.0652    -3.33  0.00087 ***
## Year2002          -0.1215      0.0625    -1.94  0.05186 .
## Year2003          -0.1708      0.0628    -2.72  0.00656 **
## Year2004          -0.2419      0.0629    -3.84  0.00012 ***
## Year2005          -0.1346      0.0607    -2.22  0.02670 *
## Year2006          -0.1279      0.0617    -2.07  0.03834 *
## Year2007          -0.1387      0.0608    -2.28  0.02259 *
## Year2008          -0.1673      0.0613    -2.73  0.00636 **
## Year2009          -0.1555      0.0588    -2.65  0.00818 **
## Year2010          -0.2139      0.0597    -3.58  0.00034 ***
## Year2011          -0.1772      0.0591    -3.00  0.00274 **
## Year2012          -0.2251      0.0612    -3.68  0.00024 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.209, Adjusted R-squared:  0.206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1021 is an outlier with |weight| = 0 ( < 2.2e-05);
## 381 weights are ~= 1. The remaining 4150 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8680 0.9450 0.9010 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.21e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1          1.014
## LastAuthorFemale 1.029 1          1.014
## Year          1.029 16          1.001

```

## Residuals from first and last author



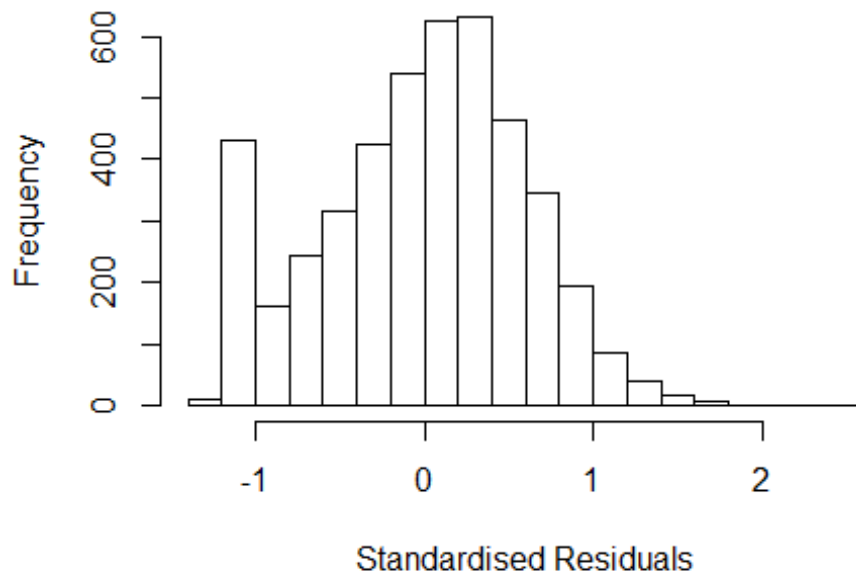
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.265 -0.417  0.048  0.399  2.383
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1688    0.0568   20.59 < 2e-16 ***
## FirstAuthorFemale1 -0.0179    0.0202   -0.89  0.375
## LastAuthorFemale1 -0.1927    0.0257   -7.48 8.7e-14 ***
## Year1997          0.0965    0.0840    1.15  0.251
## Year1998         -0.0855    0.0793   -1.08  0.281
## Year1999          0.0184    0.0732    0.25  0.801
## Year2000         -0.0503    0.0741   -0.68  0.497
## Year2001         -0.0779    0.0714   -1.09  0.276
## Year2002          0.0113    0.0679    0.17  0.868
## Year2003         -0.0609    0.0720   -0.85  0.398
## Year2004         -0.1310    0.0691   -1.90  0.058 .
## Year2005         -0.0404    0.0663   -0.61  0.543
```

```

## Year2006          0.0183      0.0668      0.27      0.784
## Year2007          0.0181      0.0662      0.27      0.785
## Year2008         -0.0300      0.0649     -0.46      0.644
## Year2009          0.0293      0.0639      0.46      0.646
## Year2010         -0.0264      0.0643     -0.41      0.681
## Year2011          0.0147      0.0648      0.23      0.821
## Year2012         -0.0164      0.0666     -0.25      0.806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0224, Adjusted R-squared:  0.0185
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 343 weights are ~= 1. The remaining 4189 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0711 0.8590 0.9490 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.026 1      1.013
## Year              1.026 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2403 -0.4124 0.0501 0.4047 2.4082
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1430 0.0580 19.72 <2e-16 ***
## FirstAuthorFemale1 -0.0547 0.0209 -2.61 0.009 **
## Year1997 0.0973 0.0862 1.13 0.259
## Year1998 -0.0871 0.0819 -1.06 0.288
## Year1999 0.0252 0.0749 0.34 0.736
## Year2000 -0.0384 0.0753 -0.51 0.610
## Year2001 -0.0772 0.0728 -1.06 0.289
## Year2002 0.0143 0.0689 0.21 0.835
## Year2003 -0.0577 0.0733 -0.79 0.431
## Year2004 -0.1282 0.0705 -1.82 0.069 .
## Year2005 -0.0297 0.0675 -0.44 0.659
## Year2006 0.0195 0.0680 0.29 0.775
```

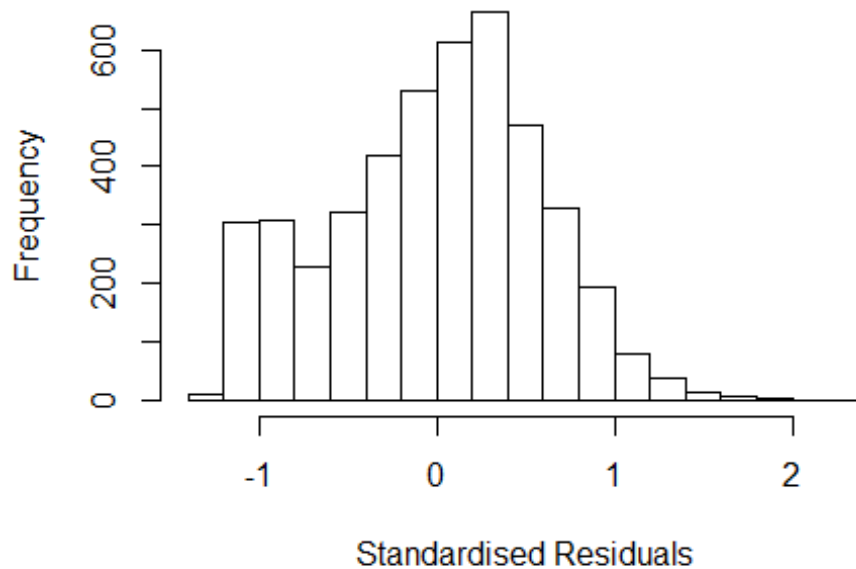
```

## Year2007          0.0178      0.0675      0.26      0.792
## Year2008         -0.0341      0.0661     -0.52      0.606
## Year2009          0.0312      0.0651      0.48      0.632
## Year2010         -0.0311      0.0655     -0.47      0.635
## Year2011          0.0181      0.0662      0.27      0.784
## Year2012         -0.0227      0.0675     -0.34      0.737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.00782,    Adjusted R-squared:  0.00408
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 349 weights are ~= 1. The remaining 4183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0652 0.8580 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1      1.008
## Year              1.017 16      1.001

```



## Residuals from last author



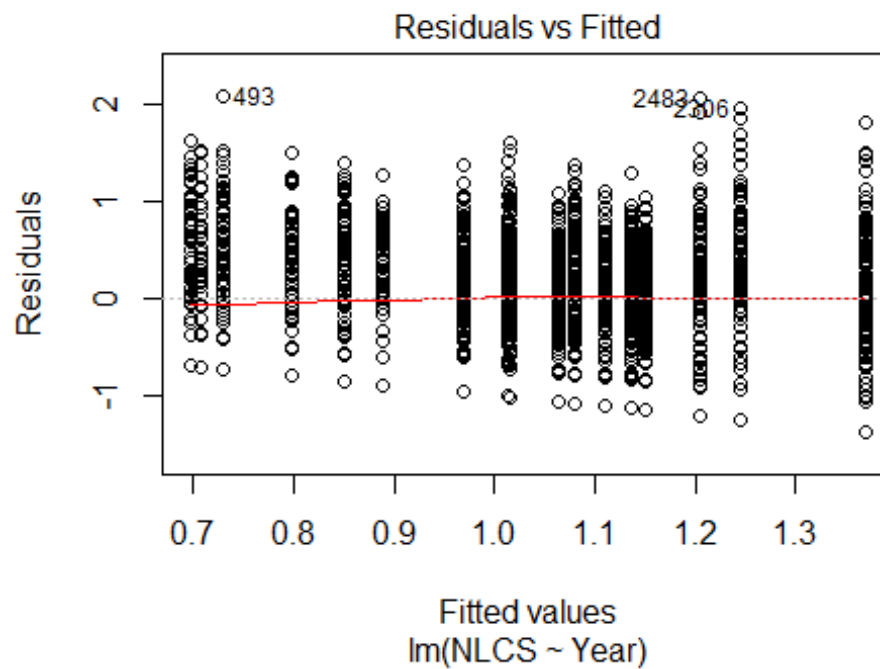
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2619 -0.4165 0.0471 0.3979 2.3870
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1648 0.0566 20.58 < 2e-16 ***
## LastAuthorFemale1 -0.1977 0.0258 -7.65 2.4e-14 ***
## Year1997 0.0971 0.0841 1.15 0.249
## Year1998 -0.0853 0.0794 -1.07 0.283
## Year1999 0.0176 0.0733 0.24 0.810
## Year2000 -0.0510 0.0740 -0.69 0.491
## Year2001 -0.0778 0.0714 -1.09 0.276
## Year2002 0.0114 0.0679 0.17 0.867
## Year2003 -0.0608 0.0721 -0.84 0.399
## Year2004 -0.1316 0.0691 -1.91 0.057 .
## Year2005 -0.0422 0.0663 -0.64 0.524
## Year2006 0.0175 0.0668 0.26 0.794
```

```

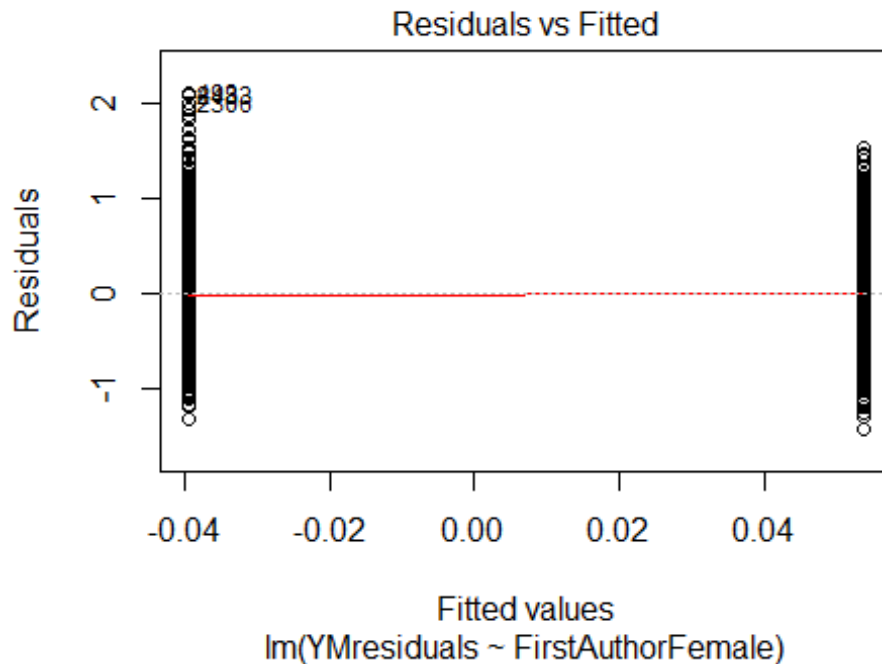
## Year2007          0.0181      0.0662      0.27      0.785
## Year2008          -0.0316     0.0649     -0.49     0.627
## Year2009          0.0274      0.0639      0.43      0.668
## Year2010          -0.0270     0.0642     -0.42     0.674
## Year2011          0.0136      0.0647      0.21      0.834
## Year2012          -0.0181     0.0665     -0.27     0.786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.594
## Multiple R-squared:  0.0222, Adjusted R-squared:  0.0185
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 350 weights are ~= 1. The remaining 4182 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.070  0.859  0.949  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4532"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2742"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 194 232 229 218 217 242 241 231 187 213 222 254 252 270 239
## 2011 2012
## 255 295
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 86 119 133 131 104 88 141 151 123 151 161 185 180 186 154
## 2011 2012

```

```
## 184 216
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 83 111 122 109 88 81 126 119 94 117 130 159 140 154 142
## 2011 2012
## 169 193
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 110, df = 16, p-value = 9e-16
```

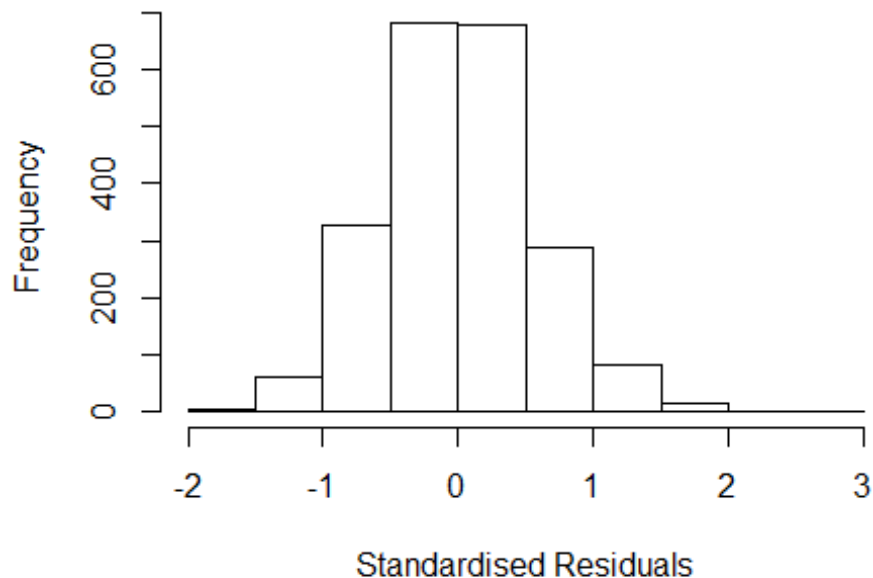


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.83, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 3.47411335439862"
## [1] "Male first author team size 2018 geometric mean: 2.96670571628752"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5700, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.41509552818623"
## [1] "Male last author team size 2018 geometric mean: 3.0955934536353"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.157 1      1.076
## LastAuthorFemale  1.188 1      1.090
## UniqueAuthors     1.443 4      1.047
## Year              1.543 16     1.014
```

## Residuals from first and last author and team size



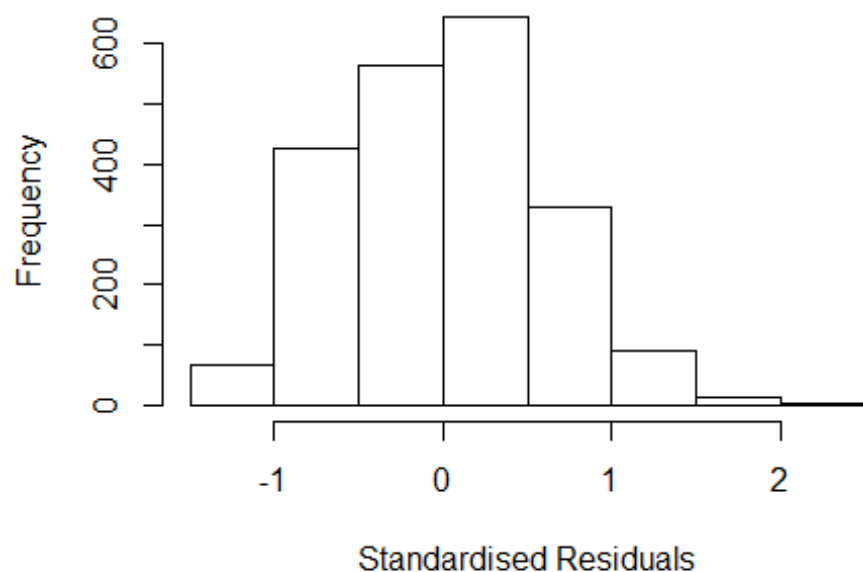
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2483 33745231692 3.263 2006      2742      1      2.613
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.56094 -0.39391 -0.00323  0.38233  2.61319
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.60168    0.06843   8.79 < 2e-16 ***
## FirstAuthorFemale1 -0.09216    0.02705  -3.41 0.00067 ***
## LastAuthorFemale1 -0.11330    0.02666  -4.25 2.2e-05 ***
## UniqueAuthors2     0.34585    0.03772   9.17 < 2e-16 ***
## UniqueAuthors3     0.49502    0.03851  12.85 < 2e-16 ***
## UniqueAuthors4     0.55748    0.04184  13.32 < 2e-16 ***
## UniqueAuthors5     0.58895    0.04387  13.43 < 2e-16 ***
## Year1997        -0.00176    0.08483  -0.02 0.98349
## Year1998        -0.05055    0.08801  -0.57 0.56579
## Year1999        -0.04799    0.08618  -0.56 0.57766
```

```

## Year2000      0.02830      0.09383      0.30  0.76300
## Year2001     -0.02502      0.08579     -0.29  0.77058
## Year2002      0.26254      0.08636      3.04  0.00239 **
## Year2003      0.19372      0.08215      2.36  0.01846 *
## Year2004      0.46852      0.11155      4.20  2.8e-05 ***
## Year2005      0.37030      0.10162      3.64  0.00027 ***
## Year2006      0.25359      0.09187      2.76  0.00582 **
## Year2007      0.20432      0.07720      2.65  0.00819 **
## Year2008      0.24284      0.07424      3.27  0.00109 **
## Year2009      0.23198      0.07660      3.03  0.00249 **
## Year2010      0.22020      0.07742      2.84  0.00450 **
## Year2011      0.09742      0.07817      1.25  0.21283
## Year2012      0.04941      0.07806      0.63  0.52681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.237, Adjusted R-squared:  0.229
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1089 is an outlier with |weight| = 0 ( < 4.7e-05);
## 168 weights are ~= 1. The remaining 1968 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.113  0.878  0.951  0.906  0.983  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           4.68e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.134 1 1.065
## LastAuthorFemale 1.159 1 1.076
## Year 1.111 16 1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41792 -0.45943 0.00713 0.42264 2.25886
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.801230 0.074222 10.79 < 2e-16 ***
## FirstAuthorFemale1 -0.077903 0.029150 -2.67 0.00759 **
## LastAuthorFemale1 -0.181402 0.028461 -6.37 2.3e-10 ***
## Year1997 0.001207 0.095109 0.01 0.98988
## Year1998 0.000217 0.095593 0.00 0.99819
## Year1999 0.016431 0.093274 0.18 0.86019
## Year2000 0.143042 0.099988 1.43 0.15269
## Year2001 0.094594 0.093506 1.01 0.31183
## Year2002 0.427903 0.096183 4.45 9.1e-06 ***
## Year2003 0.310051 0.090707 3.42 0.00064 ***
## Year2004 0.616686 0.115665 5.33 1.1e-07 ***
## Year2005 0.494509 0.106092 4.66 3.3e-06 ***
```

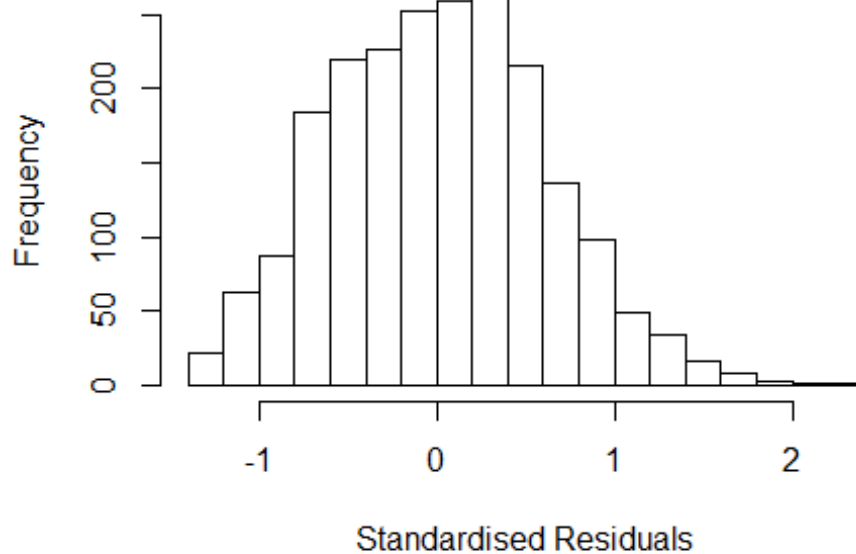
```

## Year2006          0.496261    0.092805    5.35  9.9e-08 ***
## Year2007          0.388548    0.081017    4.80  1.7e-06 ***
## Year2008          0.479718    0.077765    6.17  8.2e-10 ***
## Year2009          0.452033    0.080743    5.60  2.4e-08 ***
## Year2010          0.434338    0.081716    5.32  1.2e-07 ***
## Year2011          0.311033    0.081348    3.82  0.00014 ***
## Year2012          0.295130    0.081204    3.63  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.119, Adjusted R-squared:  0.112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1969 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.124  0.872  0.944  0.909  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.68e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.041 1      1.020
## Year              1.041 16      1.001

```



## Residuals from first author



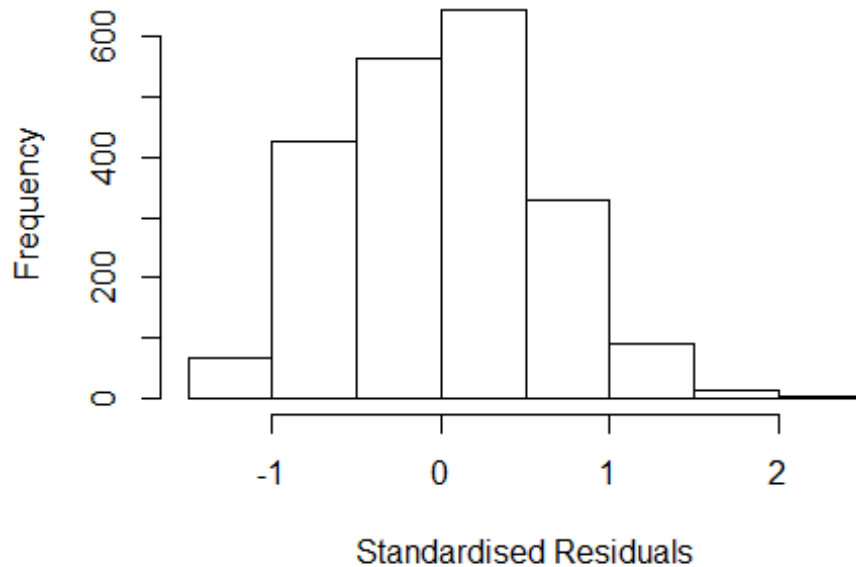
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3731 -0.4417 0.0116 0.4207 2.2056
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.72786 0.07553 9.64 < 2e-16 ***
## FirstAuthorFemale1 -0.14996 0.02813 -5.33 1.1e-07 ***
## Year1997 0.00888 0.09794 0.09 0.92777
## Year1998 0.01754 0.09930 0.18 0.85982
## Year1999 0.04014 0.09653 0.42 0.67755
## Year2000 0.15585 0.10289 1.51 0.12996
## Year2001 0.10694 0.09840 1.09 0.27726
## Year2002 0.45507 0.09953 4.57 5.1e-06 ***
## Year2003 0.31255 0.09412 3.32 0.00091 ***
## Year2004 0.64528 0.11483 5.62 2.2e-08 ***
## Year2005 0.52001 0.11091 4.69 2.9e-06 ***
## Year2006 0.52501 0.09565 5.49 4.5e-08 ***
```

```

## Year2007          0.41864    0.08328    5.03  5.4e-07 ***
## Year2008          0.51400    0.08099    6.35  2.7e-10 ***
## Year2009          0.47944    0.08337    5.75  1.0e-08 ***
## Year2010          0.47010    0.08353    5.63  2.1e-08 ***
## Year2011          0.32742    0.08402    3.90  0.00010 ***
## Year2012          0.32339    0.08394    3.85  0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0968
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 173 weights are ~= 1. The remaining 1964 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.149  0.869  0.944  0.907  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.68e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.066 1      1.032
## Year              1.066 16      1.002

```

## Residuals from last author



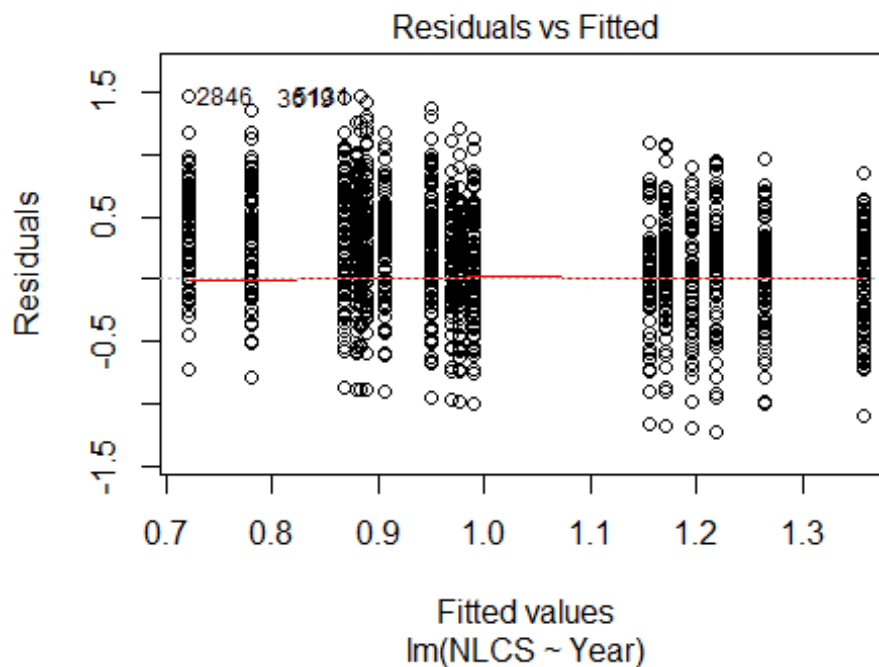
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38959 -0.46621 0.00871 0.42211 2.24127
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.76783 0.07293 10.53 < 2e-16 ***
## LastAuthorFemale1 -0.21148 0.02755 -7.68 2.5e-14 ***
## Year1997 0.00554 0.09581 0.06 0.95388
## Year1998 0.00338 0.09540 0.04 0.97172
## Year1999 0.02104 0.09406 0.22 0.82302
## Year2000 0.15208 0.10137 1.50 0.13368
## Year2001 0.08879 0.09470 0.94 0.34859
## Year2002 0.42849 0.09686 4.42 1.0e-05 ***
## Year2003 0.31375 0.09148 3.43 0.00062 ***
## Year2004 0.62177 0.11672 5.33 1.1e-07 ***
## Year2005 0.49475 0.10643 4.65 3.5e-06 ***
## Year2006 0.49664 0.09328 5.32 1.1e-07 ***
```

```

## Year2007          0.38899      0.08181      4.75  2.1e-06 ***
## Year2008          0.47747      0.07844      6.09  1.4e-09 ***
## Year2009          0.45424      0.08137      5.58  2.7e-08 ***
## Year2010          0.43413      0.08223      5.28  1.4e-07 ***
## Year2011          0.32175      0.08146      3.95  8.1e-05 ***
## Year2012          0.30068      0.08153      3.69  0.00023 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.6
## Multiple R-squared:  0.116, Adjusted R-squared:  0.109
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 162 weights are ~= 1. The remaining 1975 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.132  0.874  0.944  0.909  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.68e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2137"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2743"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  301  239  235  229  252  309  284  256  243  265  254  263  314  279  285
## 2011 2012
##  251  212
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  111   84   79   81   61   65   97  127  131  164  157  151  193  156  171
## 2011 2012

```

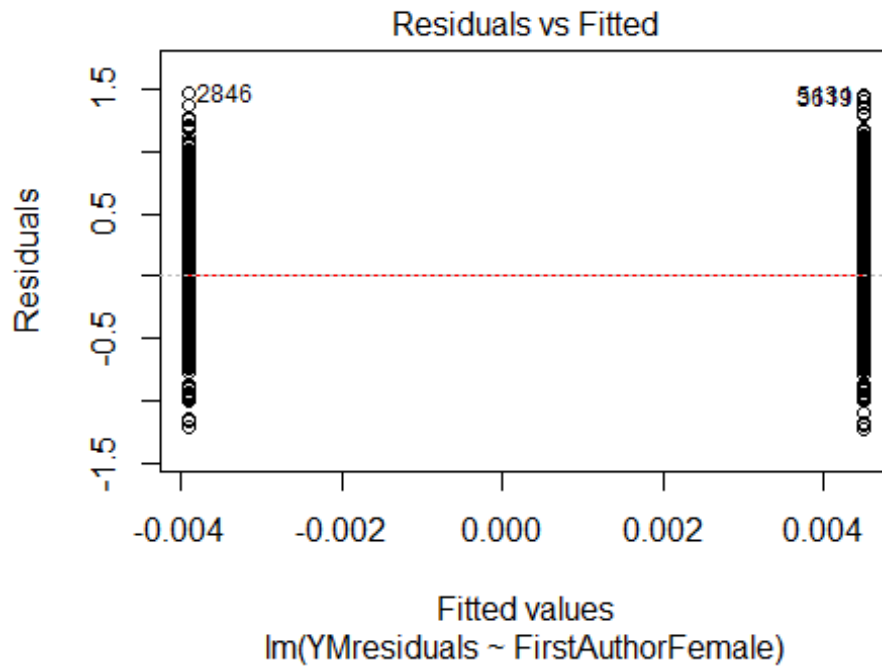
```
## 162 114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 79 62 66 55 62 76 108 97 143 136 120 152 134 151
## 2011 2012
## 142 104
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
##
## [1] "Female first author team size 2018 geometric mean: 3.07616427249092"
## [1] "Male first author team size 2018 geometric mean: 3.24070854387446"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 890, p-value = 0.6
```

```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.86306622427752"
## [1] "Male last author team size 2018 geometric mean: 3.41118411164331"

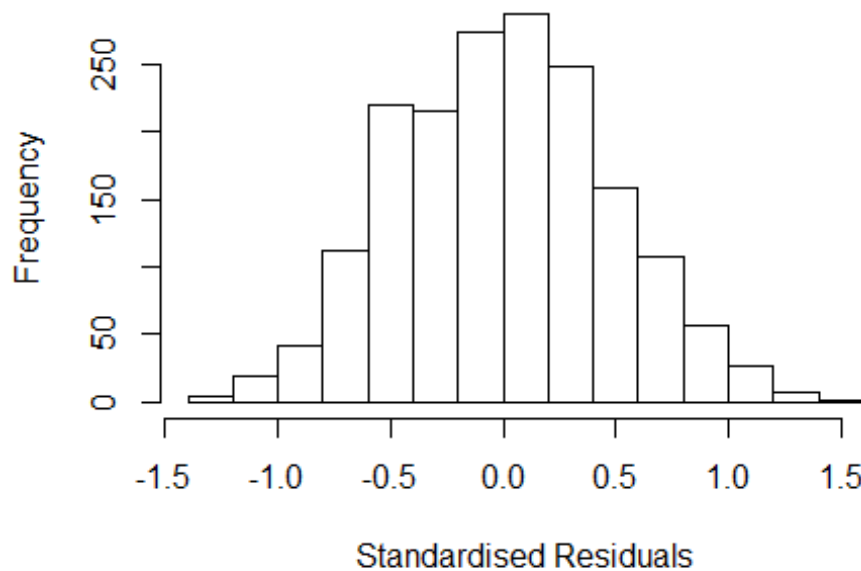
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 850, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.215	1	1.102
LastAuthorFemale	1.238	1	1.113
UniqueAuthors	1.359	4	1.039
Year	1.407	16	1.011

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.29440 -0.34528 0.00584 0.32586 1.55639
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.867968 0.063545 13.66 < 2e-16 ***
## FirstAuthorFemale1 0.007335 0.025964 0.28 0.77759
## LastAuthorFemale1 -0.000766 0.026795 -0.03 0.97719
## UniqueAuthors2 0.206837 0.040722 5.08 4.2e-07 ***
## UniqueAuthors3 0.419864 0.043316 9.69 < 2e-16 ***
## UniqueAuthors4 0.536270 0.039453 13.59 < 2e-16 ***
## UniqueAuthors5 0.613852 0.035061 17.51 < 2e-16 ***
## Year1997 -0.007010 0.083213 -0.08 0.93288
## Year1998 0.080356 0.079289 1.01 0.31098
## Year1999 -0.023410 0.073961 -0.32 0.75165
```

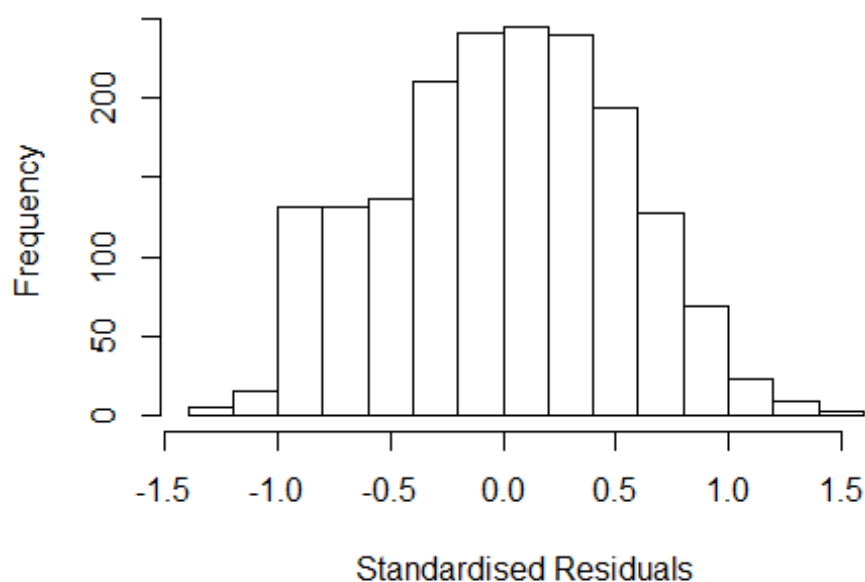
```

## Year2000      -0.147224    0.084751    -1.74    0.08254 .
## Year2001      -0.142862    0.086954    -1.64    0.10057
## Year2002      -0.300942    0.084666    -3.55    0.00039 ***
## Year2003      -0.253884    0.070664    -3.59    0.00034 ***
## Year2004      -0.269653    0.073712    -3.66    0.00026 ***
## Year2005      -0.386448    0.070477    -5.48    4.8e-08 ***
## Year2006      -0.415194    0.071873    -5.78    9.0e-09 ***
## Year2007      -0.388928    0.074100    -5.25    1.7e-07 ***
## Year2008      -0.267428    0.069958    -3.82    0.00014 ***
## Year2009      -0.331330    0.072112    -4.59    4.6e-06 ***
## Year2010      -0.343919    0.071691    -4.80    1.7e-06 ***
## Year2011      -0.321312    0.068274    -4.71    2.7e-06 ***
## Year2012      -0.364366    0.076026    -4.79    1.8e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.264, Adjusted R-squared:  0.254
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1646 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.283  0.875   0.948   0.911   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.189 1      1.090
## LastAuthorFemale  1.157 1      1.076
## Year              1.086 16      1.003

```



## Residuals from first and last author



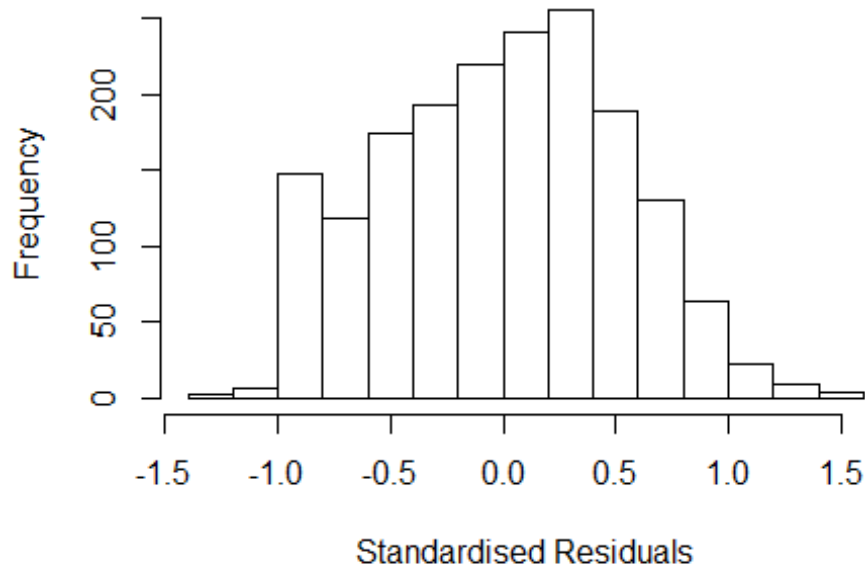
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2882 -0.3696 0.0119 0.3842 1.4733
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20617 0.05660 21.31 < 2e-16 ***
## FirstAuthorFemale1 0.03398 0.02891 1.18 0.23999
## LastAuthorFemale1 -0.10437 0.02924 -3.57 0.00037 ***
## Year1997 0.04806 0.08475 0.57 0.57070
## Year1998 0.18394 0.07219 2.55 0.01092 *
## Year1999 0.06708 0.07420 0.90 0.36607
## Year2000 -0.06266 0.08145 -0.77 0.44180
## Year2001 -0.00138 0.07991 -0.02 0.98621
## Year2002 -0.22824 0.08403 -2.72 0.00667 **
## Year2003 -0.19092 0.07133 -2.68 0.00751 **
## Year2004 -0.20531 0.07613 -2.70 0.00706 **
## Year2005 -0.49919 0.07428 -6.72 2.4e-11 ***
```

```

## Year2006      -0.42853    0.07654   -5.60  2.5e-08 ***
## Year2007      -0.37044    0.07606   -4.87  1.2e-06 ***
## Year2008      -0.23661    0.07060   -3.35  0.00082 ***
## Year2009      -0.30374    0.07591   -4.00  6.6e-05 ***
## Year2010      -0.32015    0.07694   -4.16  3.3e-05 ***
## Year2011      -0.26952    0.07044   -3.83  0.00013 ***
## Year2012      -0.30251    0.08080   -3.74  0.00019 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0924
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 124 weights are ~= 1. The remaining 1658 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.456  0.870   0.950   0.917   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.053 1      1.026
## Year      1.053 16      1.002

```

## Residuals from first author



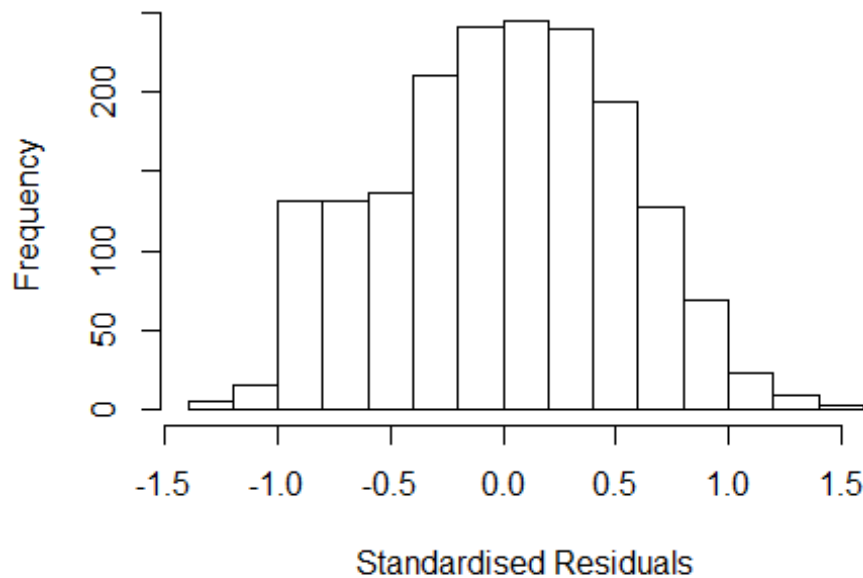
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2352 -0.4013 0.0256 0.3795 1.5030
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18488 0.05694 20.81 < 2e-16 ***
## FirstAuthorFemale1 -0.00297 0.02732 -0.11 0.91332
## Year1997 0.05027 0.08530 0.59 0.55569
## Year1998 0.18750 0.07220 2.60 0.00948 **
## Year1999 0.06893 0.07515 0.92 0.35915
## Year2000 -0.05238 0.08137 -0.64 0.51979
## Year2001 0.01130 0.08150 0.14 0.88974
## Year2002 -0.22244 0.08510 -2.61 0.00903 **
## Year2003 -0.18990 0.07196 -2.64 0.00839 **
## Year2004 -0.20414 0.07663 -2.66 0.00779 **
## Year2005 -0.50293 0.07487 -6.72 2.5e-11 ***
## Year2006 -0.42761 0.07724 -5.54 3.6e-08 ***
```

```

## Year2007          -0.37297      0.07714    -4.84  1.4e-06 ***
## Year2008          -0.24096      0.07168    -3.36  0.00079 ***
## Year2009          -0.29568      0.07645    -3.87  0.00011 ***
## Year2010          -0.31561      0.07757    -4.07  4.9e-05 ***
## Year2011          -0.26581      0.07103    -3.74  0.00019 ***
## Year2012          -0.30425      0.08157    -3.73  0.00020 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0949, Adjusted R-squared:  0.0862
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 137 weights are ~= 1. The remaining 1645 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.440  0.867  0.950  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1      1.017
## Year              1.033 16      1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2628 -0.3663 0.0208 0.3738 1.4627
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21982 0.05512 22.13 < 2e-16 ***
## LastAuthorFemale1 -0.09191 0.02761 -3.33 0.00089 ***
## Year1997 0.04298 0.08437 0.51 0.61051
## Year1998 0.18241 0.07204 2.53 0.01143 *
## Year1999 0.06854 0.07398 0.93 0.35430
## Year2000 -0.06680 0.08113 -0.82 0.41036
## Year2001 -0.00375 0.08023 -0.05 0.96272
## Year2002 -0.22634 0.08441 -2.68 0.00740 **
## Year2003 -0.19378 0.07134 -2.72 0.00667 **
## Year2004 -0.20195 0.07599 -2.66 0.00794 **
## Year2005 -0.50051 0.07398 -6.77 1.8e-11 ***
## Year2006 -0.42976 0.07640 -5.63 2.2e-08 ***
```

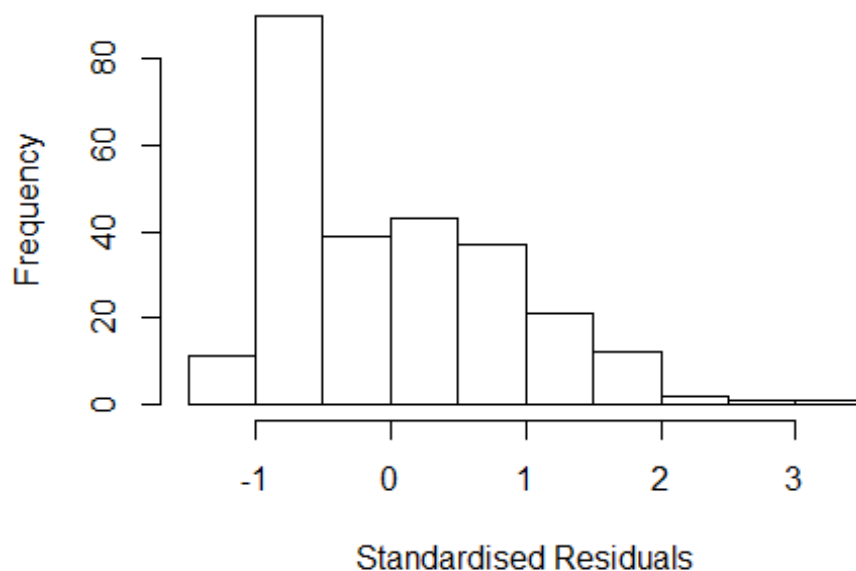
```

## Year2007          -0.36677      0.07584    -4.84  1.4e-06 ***
## Year2008          -0.23616      0.07054    -3.35  0.00083 ***
## Year2009          -0.30147      0.07587    -3.97  7.4e-05 ***
## Year2010          -0.31918      0.07667    -4.16  3.3e-05 ***
## Year2011          -0.26747      0.07025    -3.81  0.00015 ***
## Year2012          -0.29932      0.08082    -3.70  0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0922
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 128 weights are ~= 1. The remaining 1654 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   0.462  0.870  0.951  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.61e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1782"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2744"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 2006 2011
## 200 231 2 1
##
## 1996 1997 2006 2011
## 121 142 2 1
##
## 1996 1997 2006 2011
## 114 140 2 1

```

```
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -5.051e+15  1      NaN
## LastAuthorFemale  6.897e+00  1      2.626
## UniqueAuthors    -8.959e+15  4      NaN
## Year              -1.281e+15  3      NaN
```

## Residuals from first and last author and team size



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 230 0031557770 3.962 1997    2744      1    3.051
## 256 0031576106 3.484 1997    2744      1    2.861
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3055 -0.6525 -0.0139  0.6893  3.0513
```

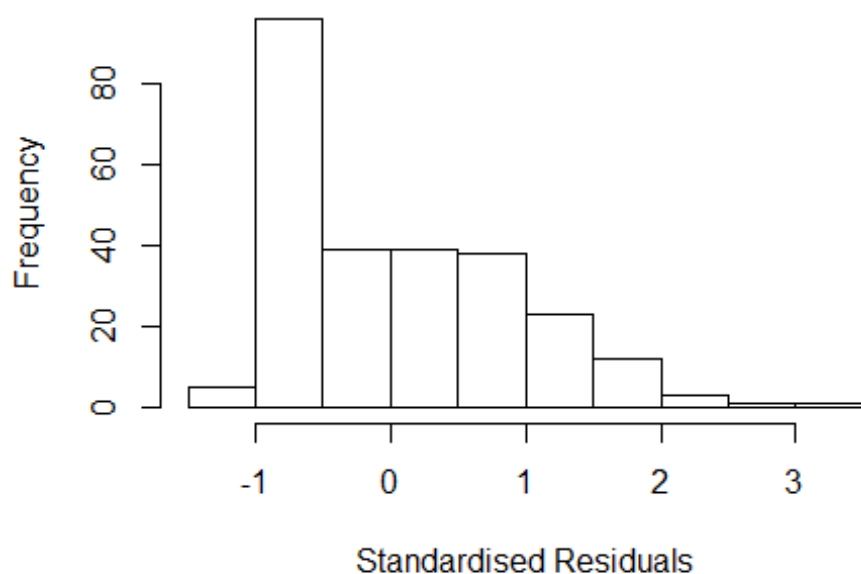
```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7148    0.1021   7.00 2.4e-11 ***
## FirstAuthorFemale1 0.3029    0.2643   1.15 0.2528
## LastAuthorFemale1 -0.2734    0.2828  -0.97 0.3346
## UniqueAuthors2    0.2878    0.1389   2.07 0.0393 *
## UniqueAuthors3   -0.1785    0.3248  -0.55 0.5832
## UniqueAuthors4   -0.2838    0.2691  -1.05 0.2927
## UniqueAuthors5   -0.3107    0.2711  -1.15 0.2529
## Year1997         -0.0919    0.1133  -0.81 0.4180
## Year2006          0.4333    0.1412   3.07 0.0024 **
## Year2011          0.8612    0.1021   8.44 2.8e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.838
## Multiple R-squared:  0.0577, Adjusted R-squared:  0.0233
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 236 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.156  0.895  0.938  0.908  0.971  0.999
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.89e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
## trace.lev    mts    compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 6.566 1          2.562
## LastAuthorFemale  3.815 1          1.953
## Year              1.260 3          1.039

```



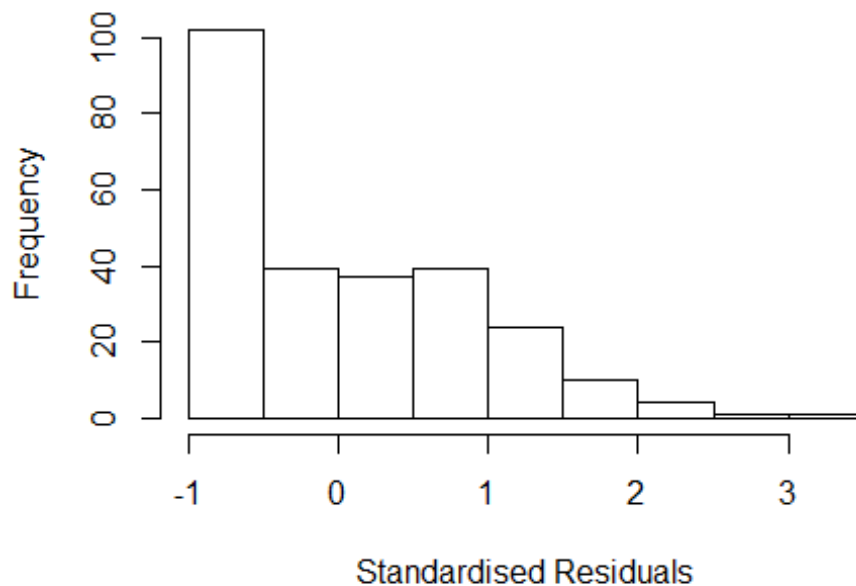
## Residuals from first and last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 230 0031557770 3.962 1997      2744      1      3.286
## 256 0031576106 3.484 1997      2744      1      2.808
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1342 -0.6761 -0.0671  0.6706  3.2859
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.786      0.098   8.02 4.0e-14 ***
## FirstAuthorFemale1  0.348      0.287   1.21  0.23
## LastAuthorFemale1 -0.336      0.303  -1.11  0.27
## Year1997          -0.110      0.115  -0.96  0.34
## Year2006           0.506      0.098   5.16 5.0e-07 ***
## Year2011           0.790      0.098   8.06 3.2e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.809
## Multiple R-squared:  0.028, Adjusted R-squared:  0.00865
```

```
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 234 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0619 0.9130 0.9350 0.8980 0.9570 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.89e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 0.3912 1      0.6254
## Year              0.3912 3      0.8552
```

## Residuals from first author

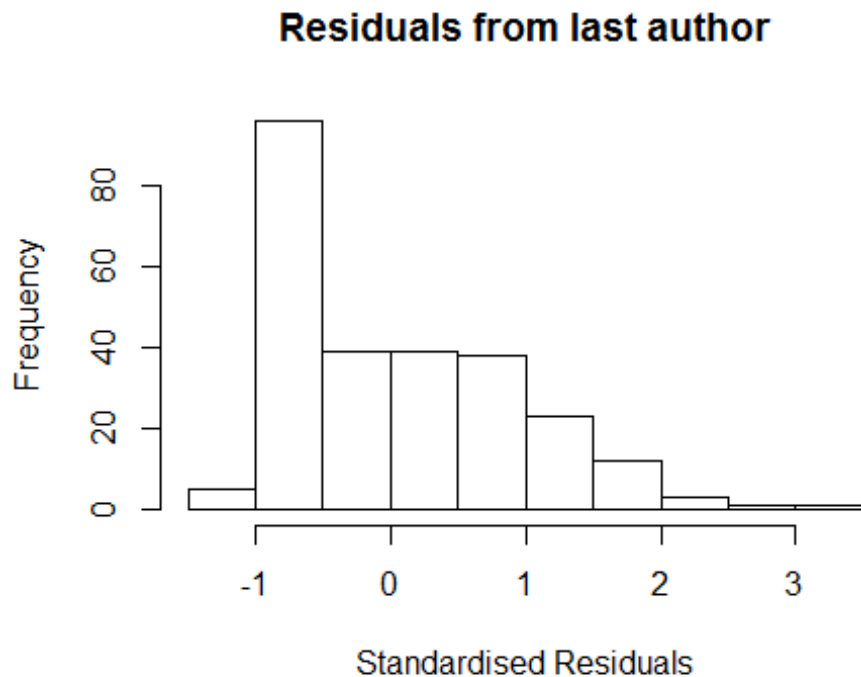


```

## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 230 0031557770 3.962 1997      2744      1      3.286
## 256 0031576106 3.484 1997      2744      1      2.808
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -0.8848 -0.6594 -0.0504  0.6522  3.3026
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.7937    0.0971   8.18 1.4e-14 ***
## FirstAuthorFemale1 0.0911    0.1302   0.70  0.48
## Year1997         -0.1343    0.1134  -1.18  0.24
## Year2006          0.4983    0.0971   5.13 5.7e-07 ***
## Year2011          0.7823    0.0971   8.06 3.1e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.805
## Multiple R-squared:  0.0167, Adjusted R-squared:  0.00107
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 17 weights are ~= 1. The remaining 240 ones are summarized as
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      0.054 0.893  0.922  0.899  0.961  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.89e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##      trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))

```

```
## LastAuthorFemale 1.855 1 1.362
## Year 1.855 3 1.108
```



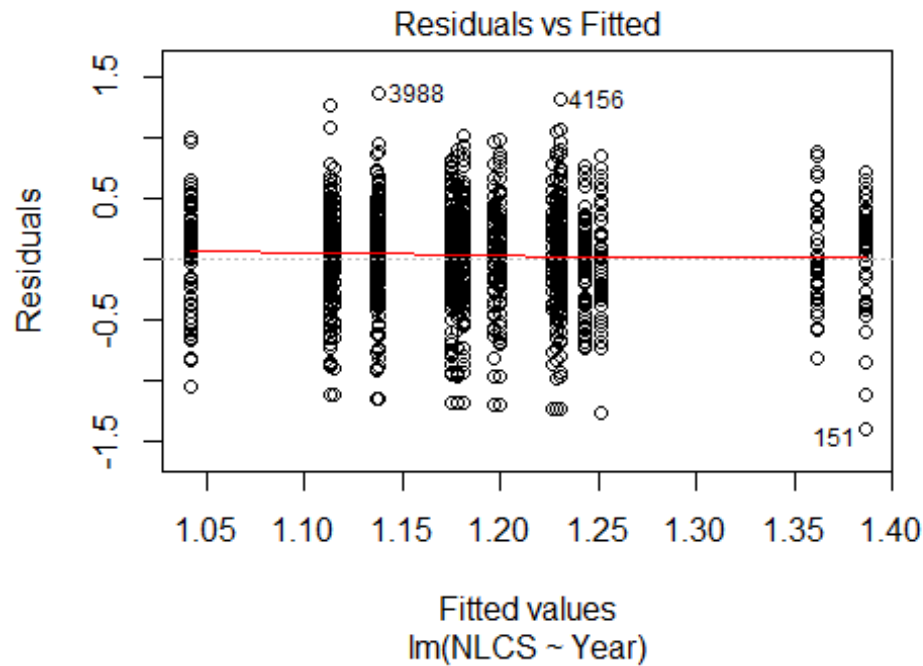
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 230 0031557770 3.962 1997    2744      1    3.286
## 256 0031576106 3.484 1997    2744      1    2.808
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8289 -0.6916 -0.0826  0.7254  3.2704
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8289    0.0981    8.45 2.4e-15 ***
## LastAuthorFemale1 -0.0323    0.1305   -0.25  0.80
## Year1997        -0.1372    0.1135   -1.21  0.23
## Year2006         0.4631    0.0981    4.72 3.9e-06 ***
## Year2011         0.7471    0.0981    7.62 5.3e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

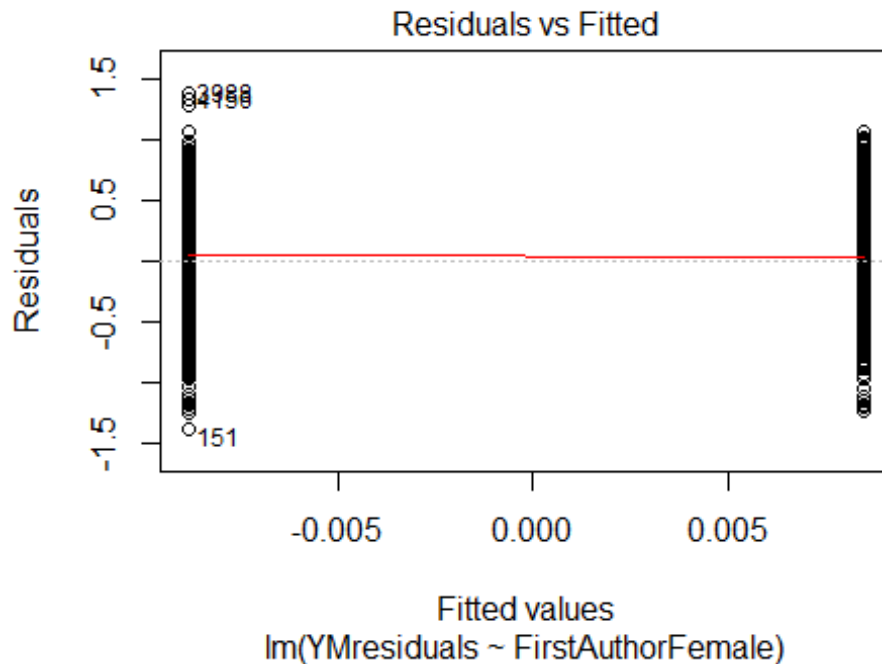
## Robust residual standard error: 0.804
## Multiple R-squared: 0.0145, Adjusted R-squared: -0.00116
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 238 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0607 0.9060 0.9340 0.8980 0.9610 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.89e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 257"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2745"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 204 185 148 151 176 215 219 190 171 192 197 261 210 202 200
## 2011 2012
## 212 212
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 36 47 58 57 42 78 76 70 91 81 155 94 123 147
## 2011 2012
## 148 145
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 43 31 43 49 49 40 68 57 50 63 53 112 63 102 128
## 2011 2012
## 135 125
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##

```

```
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 12, df = 16, p-value = 0.7
```

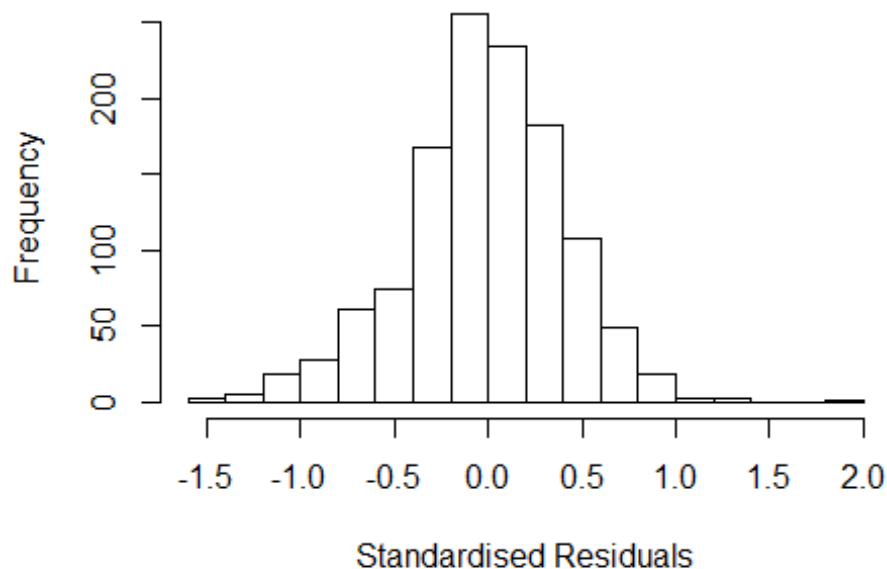


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 4.86657916868895"
## [1] "Male first author team size 2018 geometric mean: 4.95631084867767"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.86081585966489"
## [1] "Male last author team size 2018 geometric mean: 4.93842555234042"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.109 1      1.053
## LastAuthorFemale  1.133 1      1.064
## UniqueAuthors    1.347 4      1.038
## Year             1.554 16      1.014
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.40953 -0.24860 -0.00404  0.25295  1.82432
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9651    0.0843   11.44 < 2e-16 ***
## FirstAuthorFemale1  0.0201    0.0247    0.81  0.41608
## LastAuthorFemale1 -0.0449    0.0271   -1.66  0.09735 .
## UniqueAuthors2     0.2506    0.0819    3.06  0.00227 **
## UniqueAuthors3     0.4409    0.0740    5.96  3.3e-09 ***
## UniqueAuthors4     0.4988    0.0691    7.22  9.1e-13 ***
## UniqueAuthors5     0.6058    0.0640    9.47 < 2e-16 ***
## Year1997         -0.0927    0.0878   -1.06  0.29157
## Year1998         -0.1920    0.0870   -2.21  0.02747 *
## Year1999         -0.2171    0.0849   -2.56  0.01072 *
```

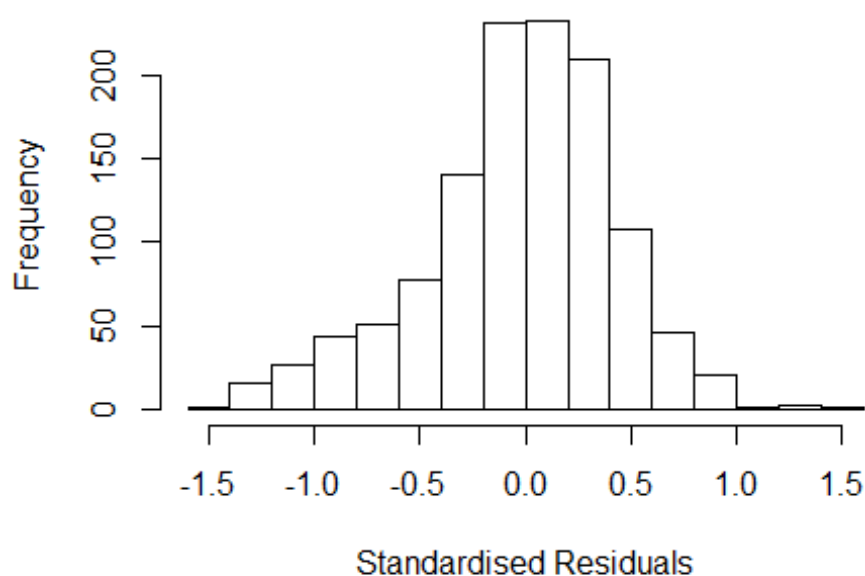


```

## Year2000          -0.2056      0.0917    -2.24   0.02513 *
## Year2001          -0.1614      0.0858    -1.88   0.06020 .
## Year2002          -0.3778      0.0860    -4.39   1.2e-05 ***
## Year2003          -0.2498      0.0781    -3.20   0.00142 **
## Year2004          -0.2514      0.0709    -3.54   0.00041 ***
## Year2005          -0.2241      0.0729    -3.07   0.00216 **
## Year2006          -0.2891      0.0768    -3.77   0.00017 ***
## Year2007          -0.3364      0.0729    -4.61   4.4e-06 ***
## Year2008          -0.2046      0.0858    -2.38   0.01729 *
## Year2009          -0.2858      0.0712    -4.02   6.3e-05 ***
## Year2010          -0.3049      0.0713    -4.28   2.0e-05 ***
## Year2011          -0.2879      0.0709    -4.06   5.2e-05 ***
## Year2012          -0.2414      0.0722    -3.34   0.00085 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.38
## Multiple R-squared:  0.184, Adjusted R-squared:  0.168
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 1155 is an outlier with |weight| = 0 ( < 8.3e-05);
## 100 weights are ~= 1. The remaining 1110 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.140  0.861  0.951  0.892  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           8.26e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1 1.032
## LastAuthorFemale 1.085 1 1.041
## Year 1.141 16 1.004

```

## Residuals from first and last author



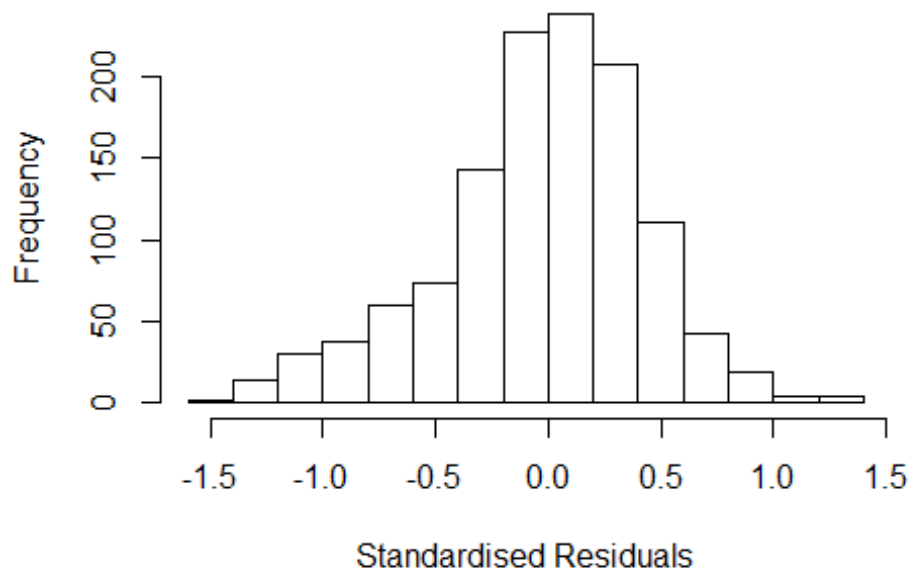
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4513 -0.2756 0.0138 0.2699 1.4003
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4513 0.0677 21.43 < 2e-16 ***
## FirstAuthorFemale1 0.0390 0.0261 1.49 0.13606
## LastAuthorFemale1 -0.0695 0.0281 -2.47 0.01368 *
## Year1997 -0.0977 0.0999 -0.98 0.32800
## Year1998 -0.1984 0.0936 -2.12 0.03417 *
## Year1999 -0.2710 0.0927 -2.92 0.00353 **
## Year2000 -0.2260 0.0941 -2.40 0.01642 *
## Year2001 -0.1860 0.0913 -2.04 0.04174 *
## Year2002 -0.3937 0.0959 -4.11 4.3e-05 ***
## Year2003 -0.2903 0.0891 -3.26 0.00116 **
## Year2004 -0.2200 0.0816 -2.70 0.00712 **
## Year2005 -0.2309 0.0840 -2.75 0.00609 **
```

```

## Year2006          -0.2426      0.0840    -2.89   0.00394 **
## Year2007          -0.2941      0.0791    -3.72   0.00021 ***
## Year2008          -0.1777      0.0894    -1.99   0.04703 *
## Year2009          -0.2748      0.0807    -3.41   0.00068 ***
## Year2010          -0.2409      0.0775    -3.11   0.00192 **
## Year2011          -0.2752      0.0771    -3.57   0.00038 ***
## Year2012          -0.1749      0.0785    -2.23   0.02607 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0348, Adjusted R-squared:  0.0202
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 108 weights are ~= 1. The remaining 1103 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.147  0.856   0.947   0.883   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1      1.029
## Year              1.059 16      1.002

```

## Residuals from first author



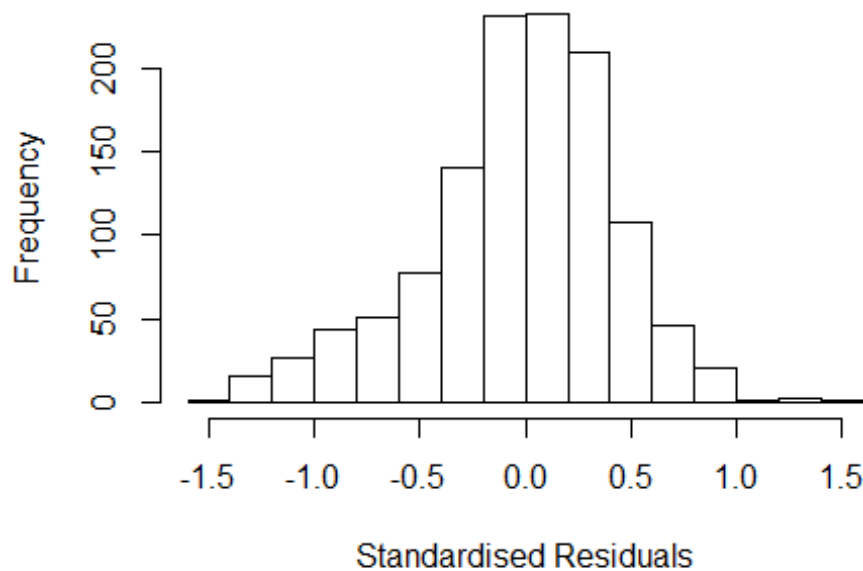
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4378 -0.2780 0.0134 0.2671 1.3538
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4378 0.0689 20.87 < 2e-16 ***
## FirstAuthorFemale1 0.0317 0.0261 1.21 0.22615
## Year1997 -0.0921 0.1013 -0.91 0.36317
## Year1998 -0.1895 0.0947 -2.00 0.04558 *
## Year1999 -0.2692 0.0935 -2.88 0.00406 **
## Year2000 -0.2221 0.0956 -2.32 0.02025 *
## Year2001 -0.1965 0.0920 -2.14 0.03293 *
## Year2002 -0.3865 0.0964 -4.01 6.5e-05 ***
## Year2003 -0.2823 0.0904 -3.12 0.00184 **
## Year2004 -0.2142 0.0832 -2.57 0.01020 *
## Year2005 -0.2327 0.0850 -2.74 0.00629 **
## Year2006 -0.2474 0.0860 -2.88 0.00408 **
```

```

## Year2007          -0.2963      0.0808    -3.67  0.00026 ***
## Year2008          -0.1848      0.0902    -2.05  0.04065 *
## Year2009          -0.2793      0.0818    -3.42  0.00066 ***
## Year2010          -0.2502      0.0789    -3.17  0.00155 **
## Year2011          -0.2845      0.0786    -3.62  0.00030 ***
## Year2012          -0.1820      0.0799    -2.28  0.02290 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0294, Adjusted R-squared:  0.0155
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 99 weights are ~= 1. The remaining 1112 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.158  0.857  0.945  0.884  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.075 1          1.037
## Year            1.075 16          1.002

```

## Residuals from last author



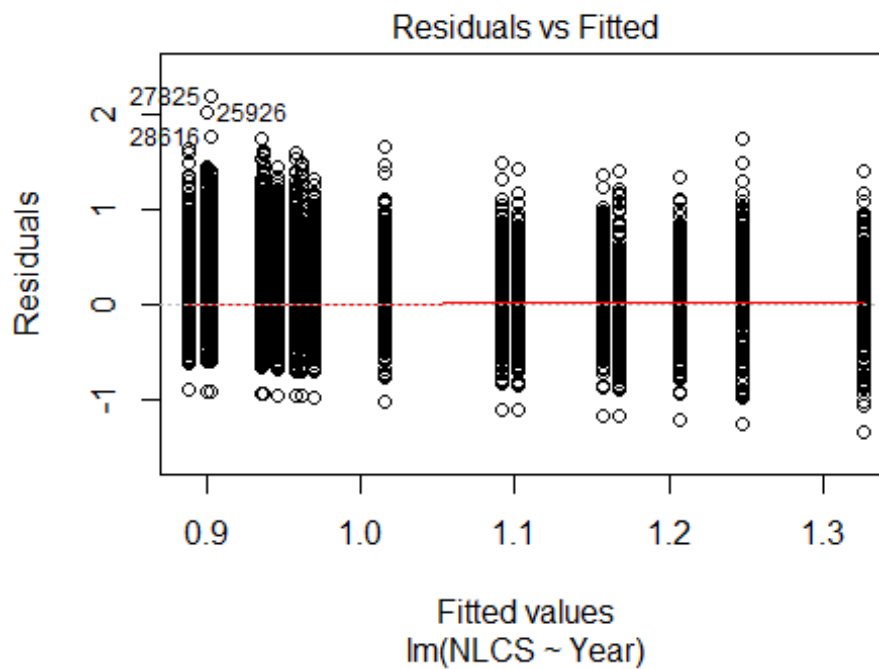
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.463 -0.274 0.017 0.268 1.376
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4633 0.0668 21.91 < 2e-16 ***
## LastAuthorFemale1 -0.0643 0.0280 -2.30 0.02184 *
## Year1997 -0.0939 0.1000 -0.94 0.34784
## Year1998 -0.1896 0.0929 -2.04 0.04149 *
## Year1999 -0.2668 0.0923 -2.89 0.00390 **
## Year2000 -0.2179 0.0926 -2.35 0.01883 *
## Year2001 -0.1839 0.0910 -2.02 0.04351 *
## Year2002 -0.3906 0.0949 -4.12 4.1e-05 ***
## Year2003 -0.2839 0.0888 -3.20 0.00142 **
## Year2004 -0.2101 0.0810 -2.59 0.00965 **
## Year2005 -0.2275 0.0835 -2.73 0.00652 **
## Year2006 -0.2334 0.0829 -2.82 0.00495 **
```

```

## Year2007          -0.2863      0.0784    -3.65   0.00027 ***
## Year2008          -0.1636      0.0880    -1.86   0.06319 .
## Year2009          -0.2653      0.0796    -3.33   0.00089 ***
## Year2010          -0.2335      0.0769    -3.04   0.00245 **
## Year2011          -0.2677      0.0765    -3.50   0.00048 ***
## Year2012          -0.1676      0.0779    -2.15   0.03171 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.033, Adjusted R-squared:  0.0192
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 115 weights are ~= 1. The remaining 1096 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.138  0.851  0.946  0.882  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1211"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2746"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1330 1222 1328 1228 1288 1434 1359 1266 1251 1468 1748 1912 1979 1928 1960
## 2011 2012
## 1923 1906
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 215 247 279 249 308 344 440 481 518 660 830 906 985 958 1070
## 2011 2012

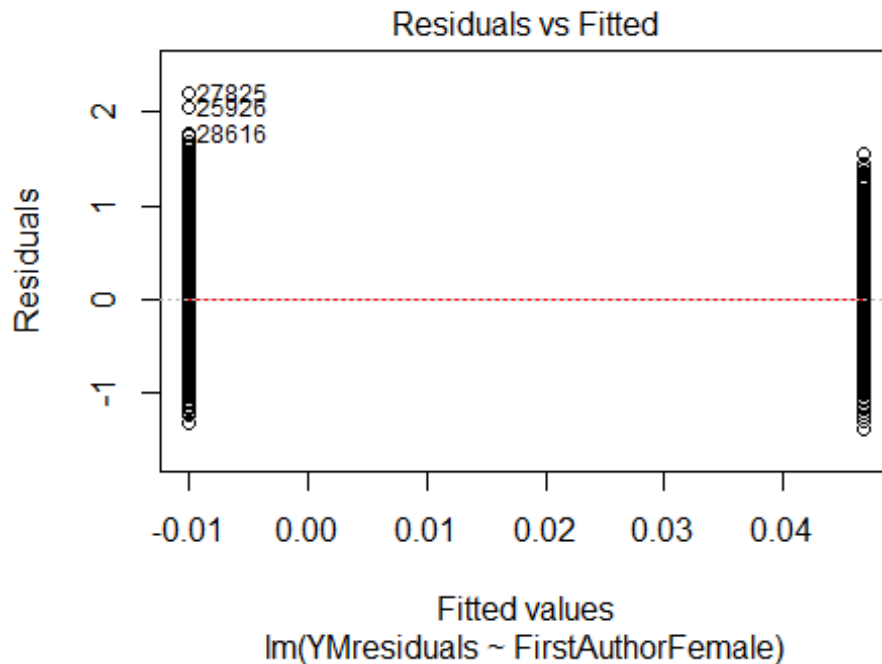
```

```
## 1059 1011
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 192 208 240 202 279 285 349 330 352 448 597 685 742 779 883
## 2011 2012
## 887 843
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 71, df = 16, p-value = 7e-09
```



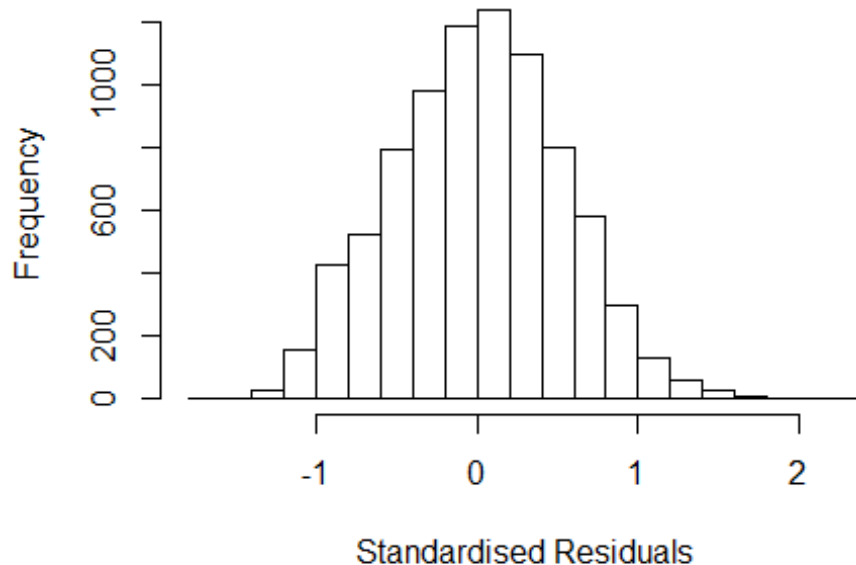
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.29, df = 1, p-value = 0.6
```





```
## [1] "Female first author team size 2018 geometric mean: 4.32105873275772"
## [1] "Male first author team size 2018 geometric mean: 4.16091721723998"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.84171769165307"
## [1] "Male last author team size 2018 geometric mean: 4.26707528919241"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## LastAuthorFemale  1.039 1          1.020
## UniqueAuthors    1.042 4          1.005
## Year              1.051 16         1.002
```

## Residuals from first and last author and team size



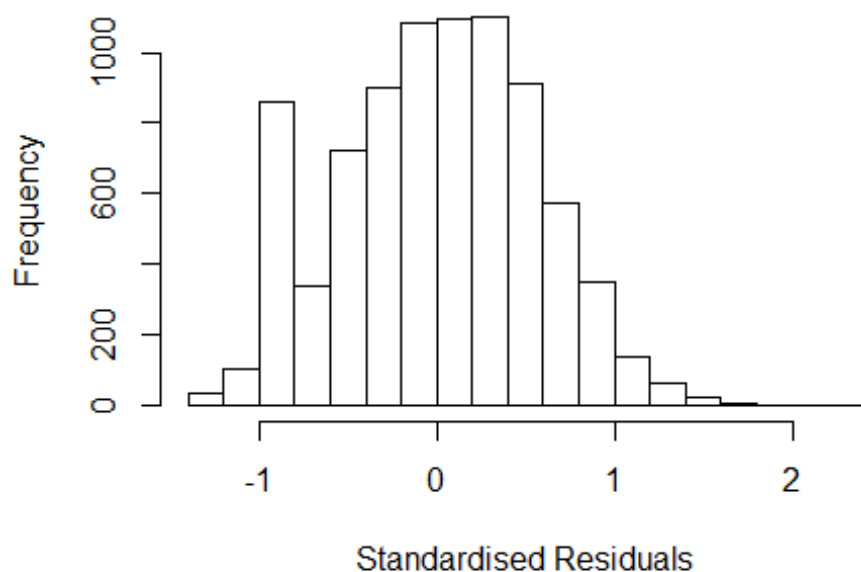
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6701 -0.3569 0.0125 0.3627 2.2130
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0092 0.0412 24.52 < 2e-16 ***
## FirstAuthorFemale1 0.0581 0.0155 3.75 0.00018 ***
## LastAuthorFemale1 0.0706 0.0204 3.46 0.00054 ***
## UniqueAuthors2 0.1589 0.0254 6.25 4.3e-10 ***
## UniqueAuthors3 0.2785 0.0244 11.40 < 2e-16 ***
## UniqueAuthors4 0.3500 0.0243 14.41 < 2e-16 ***
## UniqueAuthors5 0.6028 0.0230 26.23 < 2e-16 ***
## Year1997 -0.1661 0.0534 -3.11 0.00186 **
## Year1998 -0.0175 0.0540 -0.32 0.74561
## Year1999 -0.1486 0.0520 -2.85 0.00432 **
```

```

## Year2000          -0.1423      0.0479   -2.97   0.00299 **
## Year2001          -0.2520      0.0485   -5.20   2.1e-07 ***
## Year2002          -0.2547      0.0451   -5.65   1.7e-08 ***
## Year2003          -0.3575      0.0462   -7.74   1.1e-14 ***
## Year2004          -0.3676      0.0464   -7.92   2.7e-15 ***
## Year2005          -0.4235      0.0441   -9.61   < 2e-16 ***
## Year2006          -0.4337      0.0430  -10.08   < 2e-16 ***
## Year2007          -0.4042      0.0422   -9.57   < 2e-16 ***
## Year2008          -0.4298      0.0422  -10.19   < 2e-16 ***
## Year2009          -0.4960      0.0425  -11.66   < 2e-16 ***
## Year2010          -0.4782      0.0424  -11.28   < 2e-16 ***
## Year2011          -0.4899      0.0421  -11.62   < 2e-16 ***
## Year2012          -0.4832      0.0423  -11.41   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.16,   Adjusted R-squared:  0.158
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 705 weights are ~= 1. The remaining 7596 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0461 0.8670 0.9490 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.20e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1          1.012
## LastAuthorFemale 1.024 1          1.012
## Year              1.017 16          1.001

```

## Residuals from first and last author



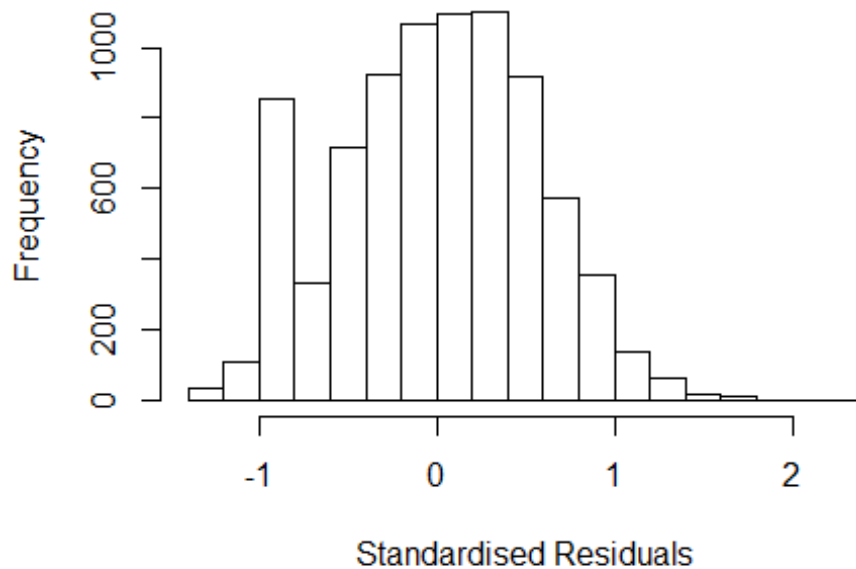
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3864 -0.3888 0.0245 0.3985 2.2055
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3265 0.0392 33.86 < 2e-16 ***
## FirstAuthorFemale1 0.0599 0.0168 3.57 0.00035 ***
## LastAuthorFemale1 0.0284 0.0224 1.27 0.20538
## Year1997 -0.1611 0.0574 -2.81 0.00502 **
## Year1998 -0.0423 0.0562 -0.75 0.45143
## Year1999 -0.1574 0.0555 -2.84 0.00456 **
## Year2000 -0.1095 0.0511 -2.14 0.03199 *
## Year2001 -0.2455 0.0517 -4.75 2.1e-06 ***
## Year2002 -0.2288 0.0479 -4.78 1.8e-06 ***
## Year2003 -0.3315 0.0489 -6.78 1.2e-11 ***
## Year2004 -0.3555 0.0487 -7.30 3.1e-13 ***
## Year2005 -0.4142 0.0461 -8.98 < 2e-16 ***
```

```

## Year2006          -0.4202      0.0458   -9.17 < 2e-16 ***
## Year2007          -0.3687      0.0445   -8.28 < 2e-16 ***
## Year2008          -0.3900      0.0446   -8.75 < 2e-16 ***
## Year2009          -0.4547      0.0445  -10.22 < 2e-16 ***
## Year2010          -0.4253      0.0448   -9.50 < 2e-16 ***
## Year2011          -0.4508      0.0445  -10.14 < 2e-16 ***
## Year2012          -0.4429      0.0450   -9.83 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.044, Adjusted R-squared:  0.0419
## Convergence in 9 IRWLS iterations
##
## Robustness weights:
## 697 weights are ~= 1. The remaining 7604 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.120  0.874  0.948  0.914  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.009 1          1.004
## Year              1.009 16          1.000

```

## Residuals from first author



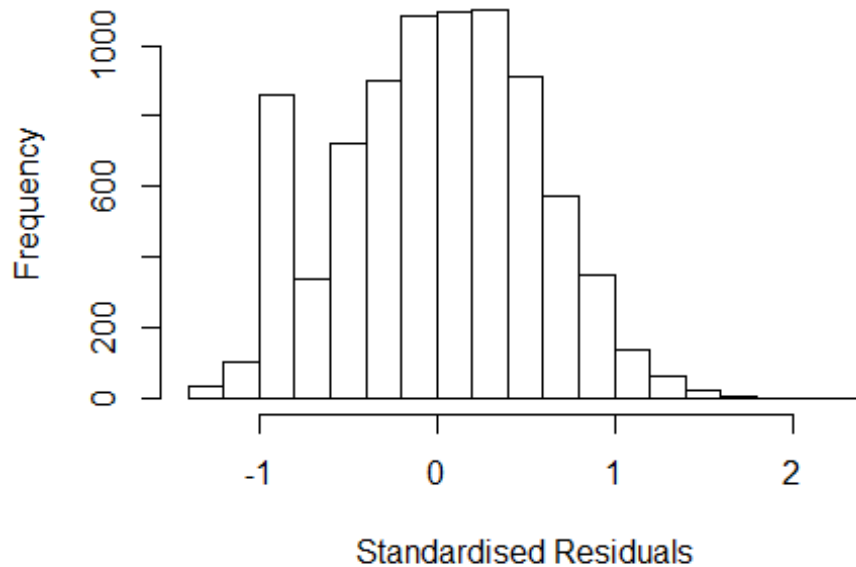
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.393 -0.389 0.024 0.400 2.203
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3288 0.0391 34.02 < 2e-16 ***
## FirstAuthorFemale1 0.0637 0.0166 3.83 0.00013 ***
## Year1997 -0.1624 0.0574 -2.83 0.00468 **
## Year1998 -0.0417 0.0561 -0.74 0.45693
## Year1999 -0.1574 0.0554 -2.84 0.00453 **
## Year2000 -0.1101 0.0510 -2.16 0.03102 *
## Year2001 -0.2457 0.0517 -4.75 2.0e-06 ***
## Year2002 -0.2292 0.0478 -4.79 1.7e-06 ***
## Year2003 -0.3323 0.0488 -6.81 1.1e-11 ***
## Year2004 -0.3564 0.0486 -7.33 2.5e-13 ***
## Year2005 -0.4147 0.0461 -9.00 < 2e-16 ***
## Year2006 -0.4202 0.0458 -9.18 < 2e-16 ***
```

```

## Year2007          -0.3689      0.0445   -8.28 < 2e-16 ***
## Year2008          -0.3898      0.0445   -8.75 < 2e-16 ***
## Year2009          -0.4548      0.0445  -10.23 < 2e-16 ***
## Year2010          -0.4253      0.0448   -9.50 < 2e-16 ***
## Year2011          -0.4511      0.0444  -10.15 < 2e-16 ***
## Year2012          -0.4427      0.0450   -9.83 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.582
## Multiple R-squared:  0.0438, Adjusted R-squared:  0.0419
## Convergence in 9 IRWLS iterations
##
## Robustness weights:
## 702 weights are ~= 1. The remaining 7599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.874  0.948  0.914  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.008 1      1.004
## Year              1.008 16      1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.334 -0.389 0.019 0.400 2.194
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3342 0.0394 33.88 < 2e-16 ***
## LastAuthorFemale1 0.0416 0.0221 1.89 0.0593 .
## Year1997 -0.1623 0.0576 -2.82 0.0049 **
## Year1998 -0.0394 0.0565 -0.70 0.4858
## Year1999 -0.1544 0.0556 -2.78 0.0055 **
## Year2000 -0.1102 0.0512 -2.15 0.0313 *
## Year2001 -0.2442 0.0518 -4.71 2.5e-06 ***
## Year2002 -0.2280 0.0480 -4.75 2.1e-06 ***
## Year2003 -0.3294 0.0491 -6.70 2.2e-11 ***
## Year2004 -0.3565 0.0489 -7.29 3.3e-13 ***
## Year2005 -0.4136 0.0464 -8.92 < 2e-16 ***
## Year2006 -0.4194 0.0460 -9.12 < 2e-16 ***
```

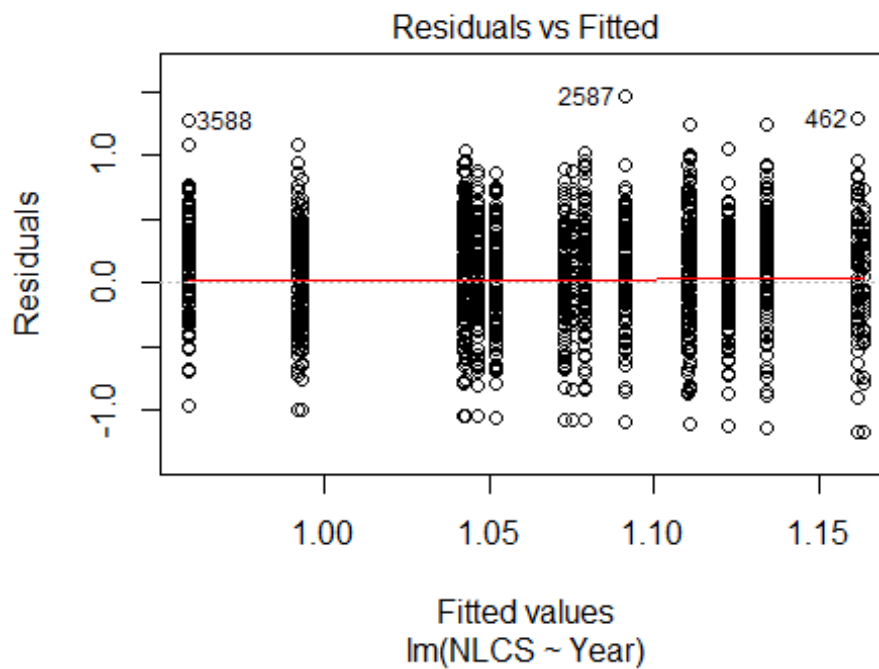


```

## Year2007          -0.3678      0.0447   -8.22  2.3e-16 ***
## Year2008          -0.3878      0.0447   -8.68  < 2e-16 ***
## Year2009          -0.4528      0.0447  -10.14  < 2e-16 ***
## Year2010          -0.4226      0.0449   -9.41  < 2e-16 ***
## Year2011          -0.4468      0.0446  -10.02  < 2e-16 ***
## Year2012          -0.4394      0.0452   -9.72  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.583
## Multiple R-squared:  0.0424, Adjusted R-squared:  0.0404
## Convergence in 9 IRWLS iterations
##
## Robustness weights:
## 698 weights are ~= 1. The remaining 7603 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.126  0.876  0.948  0.914  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.20e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 8301"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2747"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 216 184 200 196 256 259 204 205 216 201 197 191 212 189 171
## 2011 2012
## 198 155
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 54 42 84 74 112 106 119 117 124 131 104 114 142 132 114
## 2011 2012

```

```
## 149 105
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 51 36 75 60 93 93 107 88 102 95 79 100 129 114 101
## 2011 2012
## 138 87
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 50, df = 16, p-value = 2e-05
```



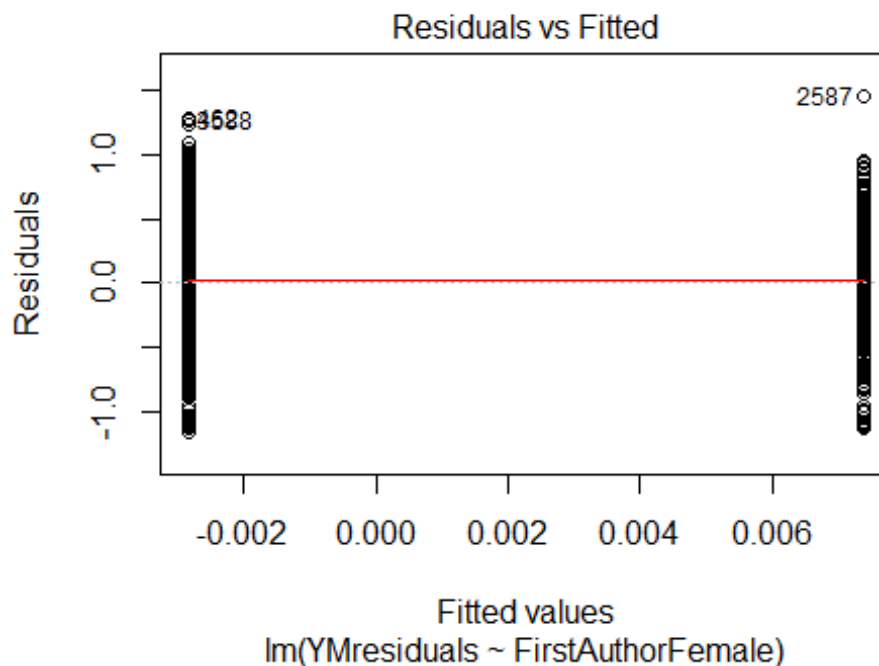
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4, df = 1, p-value = 0.05

## [1] "Female first author team size 2018 geometric mean: 5.72466412054699"
## [1] "Male first author team size 2018 geometric mean: 4.06904362840083"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

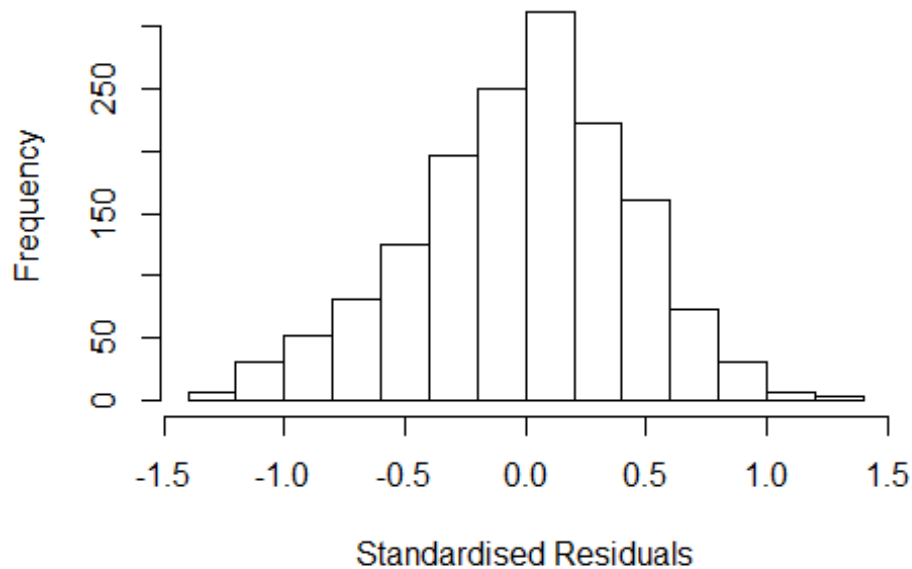
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 520, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.21693094297916"
## [1] "Male last author team size 2018 geometric mean: 4.71428394805568"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.061  1      1.030
## LastAuthorFemale  1.043  1      1.021
## UniqueAuthors    1.288  4      1.032
## Year              1.310 16      1.008
```

## Residuals from first and last author and team size



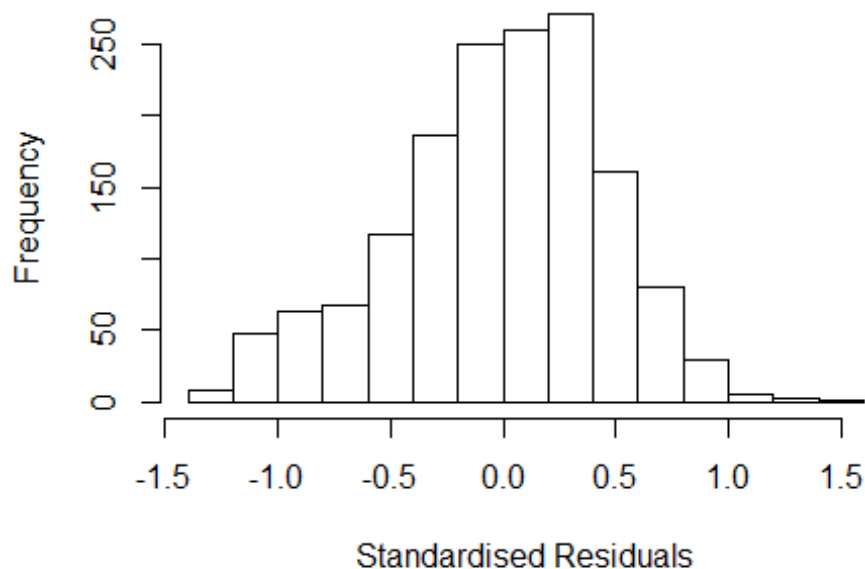
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3128 -0.3062 0.0236 0.2937 1.3798
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.807755 0.079340 10.18 < 2e-16 ***
## FirstAuthorFemale1 -0.007209 0.025203 -0.29 0.77489
## LastAuthorFemale1 0.009702 0.033810 0.29 0.77419
## UniqueAuthors2 0.043096 0.067740 0.64 0.52474
## UniqueAuthors3 0.207526 0.061719 3.36 0.00079 ***
## UniqueAuthors4 0.225253 0.057459 3.92 9.2e-05 ***
## UniqueAuthors5 0.420677 0.049336 8.53 < 2e-16 ***
## Year1997 0.119269 0.085799 1.39 0.16470
## Year1998 0.068603 0.083321 0.82 0.41043
## Year1999 -0.000589 0.077984 -0.01 0.99397
```

```

## Year2000      0.046266  0.069419  0.67  0.50521
## Year2001      0.042745  0.072272  0.59  0.55431
## Year2002      0.024745  0.075255  0.33  0.74234
## Year2003     -0.108003  0.073433 -1.47  0.14156
## Year2004     -0.060057  0.070243 -0.85  0.39269
## Year2005     -0.036329  0.074026 -0.49  0.62366
## Year2006     -0.010464  0.073573 -0.14  0.88692
## Year2007     -0.046977  0.078403 -0.60  0.54914
## Year2008      0.084334  0.072240  1.17  0.24323
## Year2009      0.025696  0.081235  0.32  0.75180
## Year2010     -0.081164  0.081900 -0.99  0.32183
## Year2011     -0.107966  0.072927 -1.48  0.13895
## Year2012     -0.169377  0.076122 -2.23  0.02622 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.122, Adjusted R-squared:  0.11
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1407 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.288  0.864  0.948  0.901  0.986  0.999
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1      1.019
## LastAuthorFemale  1.035 1      1.018
## Year              1.046 16      1.001

```

## Residuals from first and last author



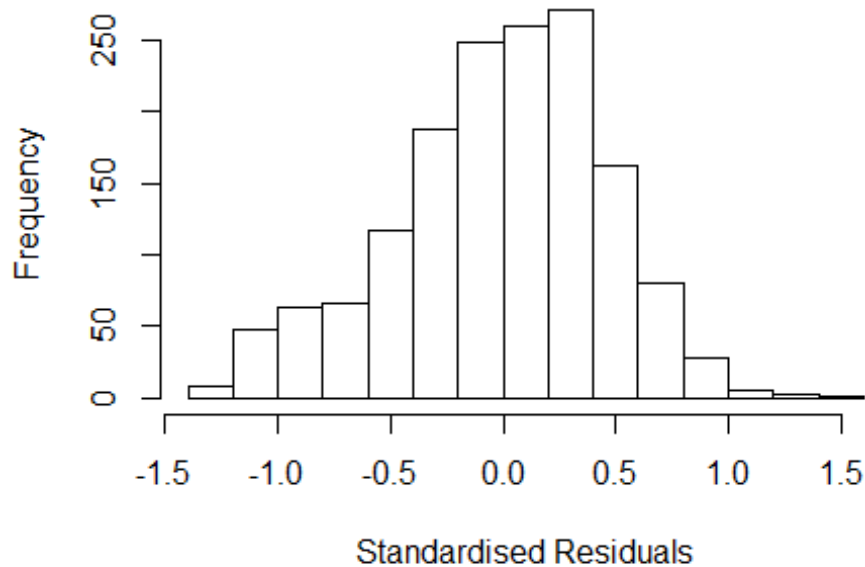
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.217 -0.303 0.030 0.313 1.452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.127215 0.061174 18.43 <2e-16 ***
## FirstAuthorFemale1 0.000013 0.026104 0.00 1.000
## LastAuthorFemale1 -0.003875 0.034640 -0.11 0.911
## Year1997 0.109047 0.083391 1.31 0.191
## Year1998 0.061508 0.085652 0.72 0.473
## Year1999 -0.048645 0.081370 -0.60 0.550
## Year2000 0.022876 0.072895 0.31 0.754
## Year2001 -0.019077 0.077374 -0.25 0.805
## Year2002 -0.027810 0.077625 -0.36 0.720
## Year2003 -0.123255 0.075203 -1.64 0.101
## Year2004 -0.066383 0.071328 -0.93 0.352
## Year2005 -0.031145 0.074277 -0.42 0.675
```

```

## Year2006          -0.026353    0.076205    -0.35    0.730
## Year2007          -0.025156    0.079315    -0.32    0.751
## Year2008           0.089489    0.075017     1.19    0.233
## Year2009           0.037382    0.082984     0.45    0.652
## Year2010          -0.076280    0.085395    -0.89    0.372
## Year2011          -0.153077    0.077217    -1.98    0.048 *
## Year2012          -0.161995    0.078936    -2.05    0.040 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.0253, Adjusted R-squared:  0.0138
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1435 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.275  0.871  0.950  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1      1.012
## Year              1.025 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2161 -0.3023 0.0309 0.3131 1.4530
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.126394 0.061159 18.42 <2e-16 ***
## FirstAuthorFemale1 -0.000375 0.025932 -0.01 0.988
## Year1997 0.109011 0.083331 1.31 0.191
## Year1998 0.061772 0.085651 0.72 0.471
## Year1999 -0.048132 0.081368 -0.59 0.554
## Year2000 0.023377 0.072989 0.32 0.749
## Year2001 -0.018555 0.077389 -0.24 0.811
## Year2002 -0.027491 0.077743 -0.35 0.724
## Year2003 -0.122864 0.075309 -1.63 0.103
## Year2004 -0.066031 0.071466 -0.92 0.356
## Year2005 -0.030764 0.074437 -0.41 0.679
## Year2006 -0.025873 0.076264 -0.34 0.734
```

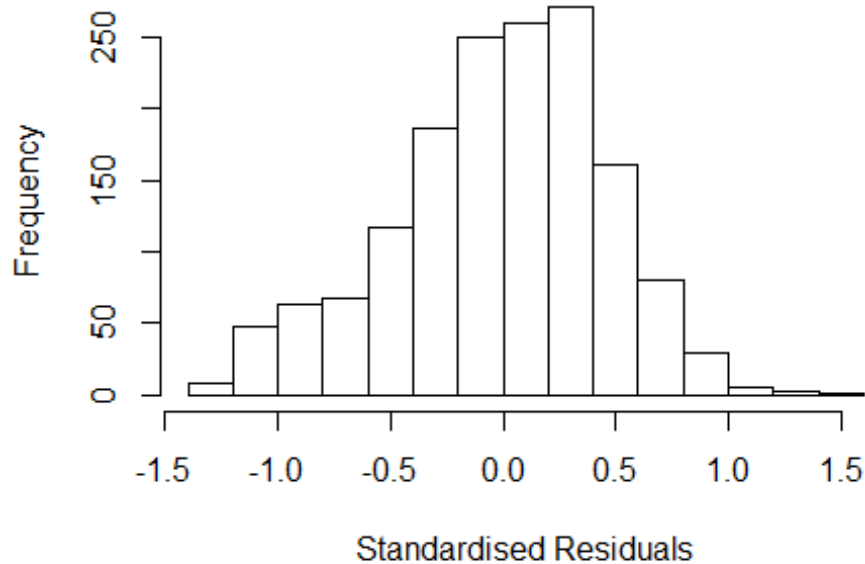


```

## Year2007          -0.025039    0.079376   -0.32    0.752
## Year2008          0.089724    0.075046    1.20    0.232
## Year2009          0.037718    0.083122    0.45    0.650
## Year2010         -0.075904    0.085392   -0.89    0.374
## Year2011         -0.152792    0.077288   -1.98    0.048 *
## Year2012         -0.161700    0.078995   -2.05    0.041 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.45
## Multiple R-squared:  0.0252, Adjusted R-squared:  0.0144
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1435 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.275  0.870  0.950  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.022 1          1.011
## Year            1.022 16          1.001

```

## Residuals from last author



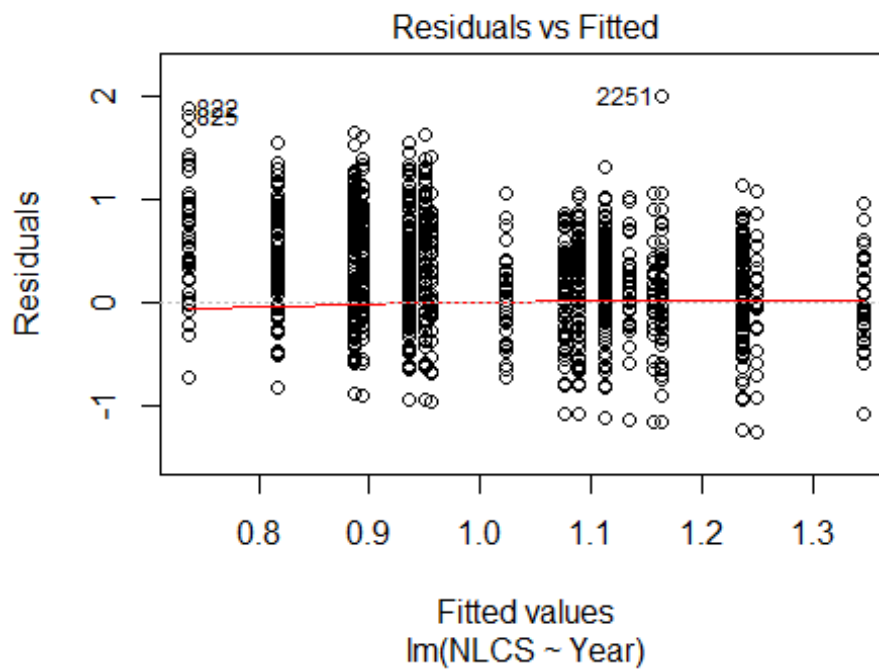
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.217 -0.303 0.030 0.313 1.452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12725 0.06040 18.66 <2e-16 ***
## LastAuthorFemale1 -0.00387 0.03442 -0.11 0.911
## Year1997 0.10901 0.08338 1.31 0.191
## Year1998 0.06150 0.08563 0.72 0.473
## Year1999 -0.04868 0.08135 -0.60 0.550
## Year2000 0.02285 0.07288 0.31 0.754
## Year2001 -0.01911 0.07735 -0.25 0.805
## Year2002 -0.02783 0.07759 -0.36 0.720
## Year2003 -0.12327 0.07519 -1.64 0.101
## Year2004 -0.06641 0.07125 -0.93 0.351
## Year2005 -0.03115 0.07429 -0.42 0.675
## Year2006 -0.02638 0.07627 -0.35 0.730
```

```

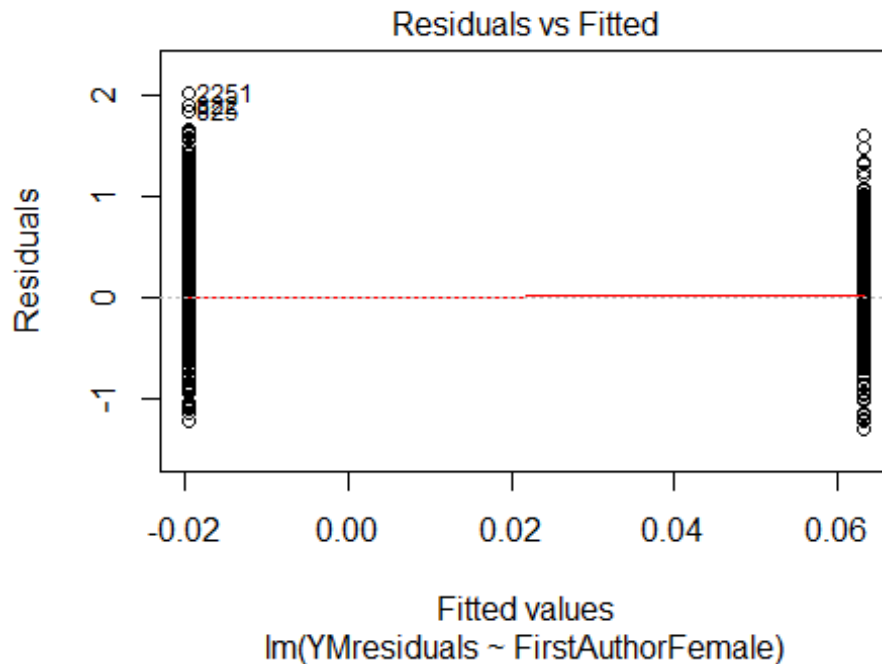
## Year2007          -0.02516      0.07931    -0.32      0.751
## Year2008           0.08951      0.07500      1.19      0.233
## Year2009           0.03740      0.08297      0.45      0.652
## Year2010          -0.07629      0.08542     -0.89      0.372
## Year2011          -0.15309      0.07724     -1.98      0.048 *
## Year2012          -0.16202      0.07895     -2.05      0.040 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.0253, Adjusted R-squared:  0.0144
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~ = 1. The remaining 1435 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.274  0.871  0.950  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.46e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1548"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2748"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  290  544  399  200  196  225  250  214  199  205  240  265  293  311  266
## 2011 2012
##  282  246
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   79   23   27   32   42   49   59   77  104  152  188  197  211  180
## 2011 2012

```

```
## 184 150
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 31 68 20 22 28 35 37 42 63 89 129 157 172 173 141
## 2011 2012
## 156 124
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 83, df = 16, p-value = 5e-11
```

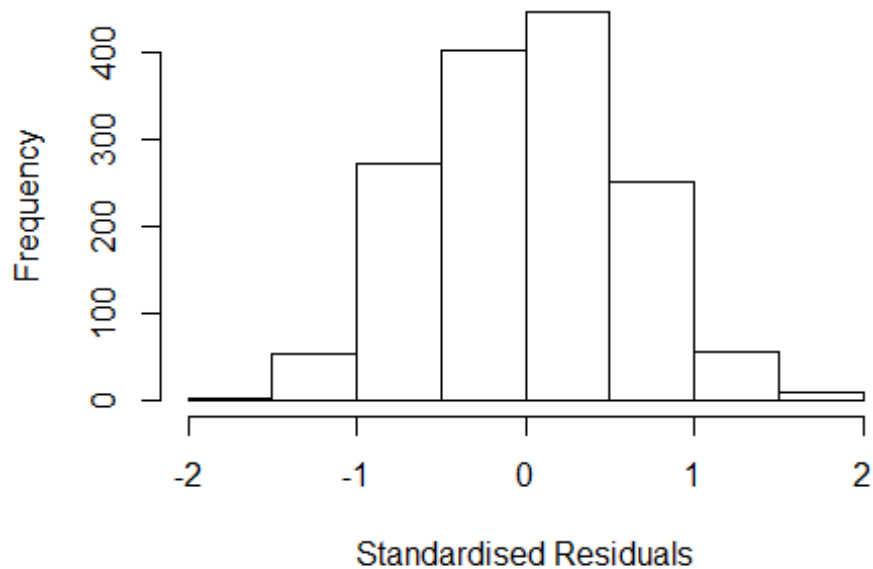


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.1, df = 1, p-value = 0.08
```



```
## [1] "Female first author team size 2018 geometric mean: 5.11892362205265"
## [1] "Male first author team size 2018 geometric mean: 4.48517367898252"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.38011110961648"
## [1] "Male last author team size 2018 geometric mean: 4.5321162916915"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.079 1      1.039
## LastAuthorFemale  1.093 1      1.046
## UniqueAuthors    1.315 4      1.035
## Year             1.390 16      1.010
```

## Residuals from first and last author and team size



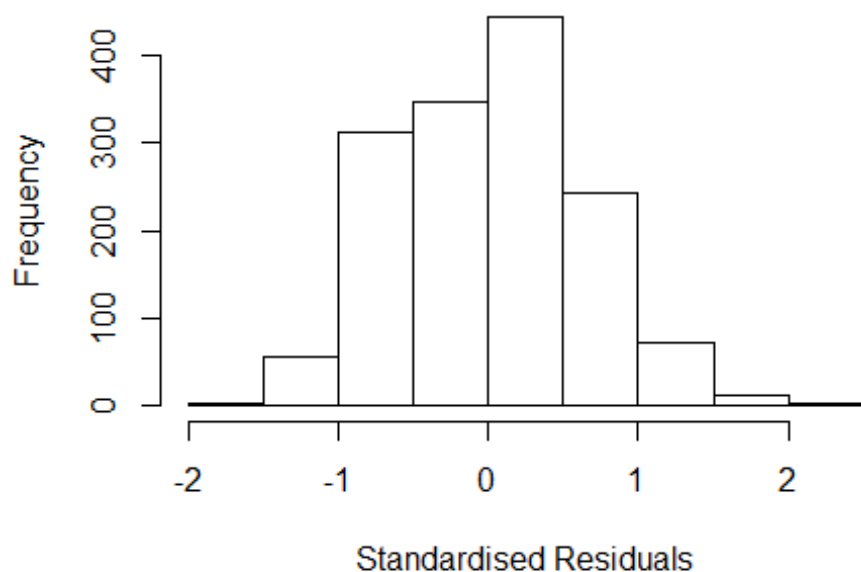
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6952 -0.4456 0.0161 0.4375 1.8973
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6587 0.1199 5.50 4.6e-08 ***
## FirstAuthorFemale1 0.0777 0.0386 2.01 0.0444 *
## LastAuthorFemale1 0.2280 0.0472 4.83 1.5e-06 ***
## UniqueAuthors2 0.3427 0.0710 4.82 1.5e-06 ***
## UniqueAuthors3 0.3887 0.0623 6.24 5.9e-10 ***
## UniqueAuthors4 0.4772 0.0628 7.60 5.2e-14 ***
## UniqueAuthors5 0.6272 0.0568 11.04 < 2e-16 ***
## Year1997 -0.3932 0.1564 -2.51 0.0120 *
## Year1998 0.1232 0.1612 0.76 0.4448
## Year1999 0.1036 0.1510 0.69 0.4930
```

```

## Year2000      0.1550      0.1435      1.08      0.2805
## Year2001     -0.1274      0.1378     -0.92      0.3554
## Year2002     -0.0252      0.1425     -0.18      0.8595
## Year2003     -0.2255      0.1524     -1.48      0.1390
## Year2004     -0.0585      0.1267     -0.46      0.6447
## Year2005      0.0262      0.1228      0.21      0.8312
## Year2006      0.0891      0.1146      0.78      0.4369
## Year2007     -0.0450      0.1144     -0.39      0.6942
## Year2008     -0.1868      0.1192     -1.57      0.1174
## Year2009     -0.3030      0.1229     -2.47      0.0138 *
## Year2010     -0.2703      0.1248     -2.17      0.0305 *
## Year2011     -0.3983      0.1238     -3.22      0.0013 **
## Year2012     -0.2635      0.1293     -2.04      0.0418 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.156, Adjusted R-squared:  0.143
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1374 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.342  0.868  0.949  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.048 1      1.024
## LastAuthorFemale  1.065 1      1.032
## Year              1.091 16      1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5167 -0.4974 0.0289 0.4438 2.0479
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04807 0.11089 9.45 < 2e-16 ***
## FirstAuthorFemale1 0.09328 0.03952 2.36 0.0184 *
## LastAuthorFemale1 0.20434 0.04991 4.09 4.5e-05 ***
## Year1997 -0.42258 0.15079 -2.80 0.0051 **
## Year1998 0.16994 0.16594 1.02 0.3059
## Year1999 0.08722 0.15875 0.55 0.5828
## Year2000 0.20784 0.13473 1.54 0.1232
## Year2001 -0.07736 0.13454 -0.58 0.5654
## Year2002 0.06200 0.14739 0.42 0.6741
## Year2003 -0.13237 0.15532 -0.85 0.3942
## Year2004 -0.00828 0.12733 -0.06 0.9482
## Year2005 0.06512 0.12803 0.51 0.6111
```

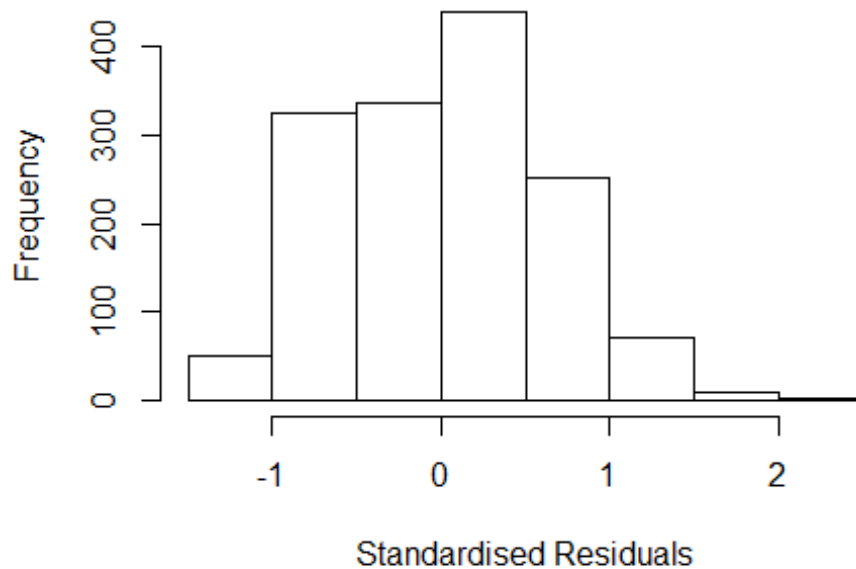


```

## Year2006          0.17099    0.11827    1.45    0.1485
## Year2007          0.03177    0.11860    0.27    0.7888
## Year2008         -0.16052    0.12536   -1.28    0.2006
## Year2009         -0.23645    0.12917   -1.83    0.0674 .
## Year2010         -0.20235    0.12690   -1.59    0.1110
## Year2011         -0.33776    0.12698   -2.66    0.0079 **
## Year2012         -0.18172    0.13292   -1.37    0.1718
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.0794, Adjusted R-squared:  0.0682
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1356 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.294  0.861  0.945  0.912  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling            cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.035 1      1.017
## Year              1.035 16      1.001

```

## Residuals from first author



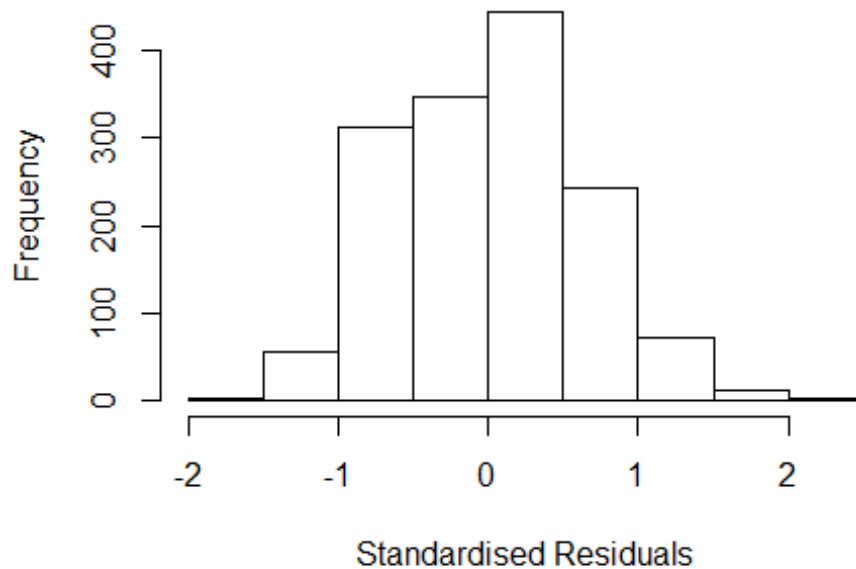
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3607 -0.5105 0.0351 0.4403 2.0231
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09091 0.10907 10.00 <2e-16 ***
## FirstAuthorFemale1 0.11951 0.03936 3.04 0.0024 **
## Year1997 -0.44562 0.15103 -2.95 0.0032 **
## Year1998 0.15433 0.16754 0.92 0.3571
## Year1999 0.04929 0.15474 0.32 0.7501
## Year2000 0.18812 0.13420 1.40 0.1612
## Year2001 -0.10235 0.13452 -0.76 0.4469
## Year2002 0.04395 0.14948 0.29 0.7688
## Year2003 -0.13068 0.15446 -0.85 0.3977
## Year2004 -0.01894 0.12657 -0.15 0.8811
## Year2005 0.04869 0.12688 0.38 0.7012
## Year2006 0.15031 0.11654 1.29 0.1973
```

```

## Year2007          0.00116    0.11692    0.01    0.9921
## Year2008          -0.19172    0.12371   -1.55    0.1214
## Year2009          -0.24646    0.12749   -1.93    0.0534 .
## Year2010          -0.22829    0.12557   -1.82    0.0693 .
## Year2011          -0.35718    0.12571   -2.84    0.0046 **
## Year2012          -0.20871    0.13058   -1.60    0.1102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.655
## Multiple R-squared:  0.0678, Adjusted R-squared:  0.057
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 130 weights are ~= 1. The remaining 1357 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.319  0.854  0.944  0.913  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.05 1      1.025
## Year              1.05 16      1.002

```

## Residuals from last author



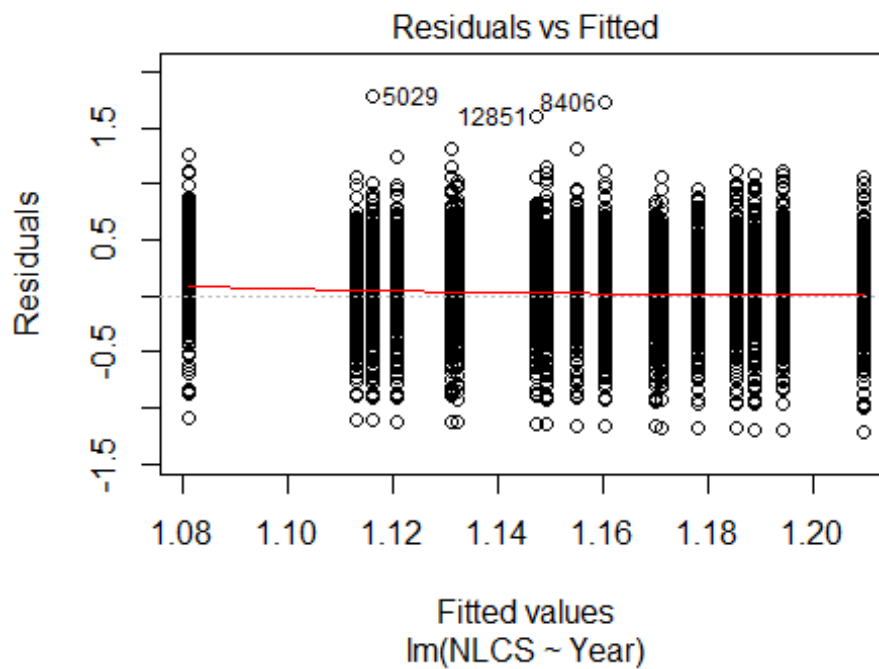
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4549 -0.5083 0.0401 0.4399 2.0285
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0640 0.1079 9.86 < 2e-16 ***
## LastAuthorFemale1 0.2233 0.0493 4.53 6.3e-06 ***
## Year1997 -0.4167 0.1495 -2.79 0.0054 **
## Year1998 0.1643 0.1662 0.99 0.3231
## Year1999 0.0877 0.1555 0.56 0.5727
## Year2000 0.2190 0.1325 1.65 0.0987 .
## Year2001 -0.0775 0.1316 -0.59 0.5558
## Year2002 0.0655 0.1443 0.45 0.6498
## Year2003 -0.1217 0.1520 -0.80 0.4235
## Year2004 -0.0071 0.1250 -0.06 0.9547
## Year2005 0.0660 0.1254 0.53 0.5985
## Year2006 0.1677 0.1155 1.45 0.1468
```

```

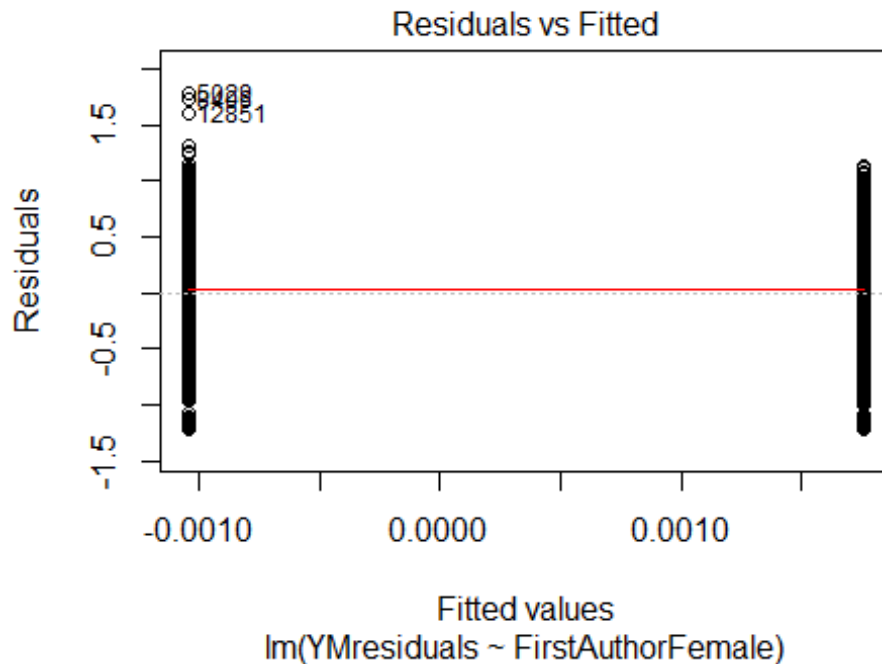
## Year2007          0.0376      0.1156      0.33      0.7448
## Year2008          -0.1534      0.1227     -1.25      0.2114
## Year2009          -0.2291      0.1263     -1.81      0.0700 .
## Year2010          -0.1931      0.1245     -1.55      0.1211
## Year2011          -0.3297      0.1241     -2.66      0.0080 **
## Year2012          -0.1755      0.1307     -1.34      0.1794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.646
## Multiple R-squared:  0.0758, Adjusted R-squared:  0.0651
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 138 weights are ~= 1. The remaining 1349 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.303  0.854  0.941  0.911  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.72e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1487"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2800"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  666  664  617  532  629  681  626  599  591  579  640  614  551  407  452
## 2011 2012
##  456  514
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  261  264  280  283  298  304  367  372  387  384  415  421  370  290  314
## 2011 2012

```

```
## 341 405
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 230 244 246 246 270 281 327 319 333 320 365 365 327 258 274
## 2011 2012
## 308 375
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 110, df = 16, p-value = 4e-16
```

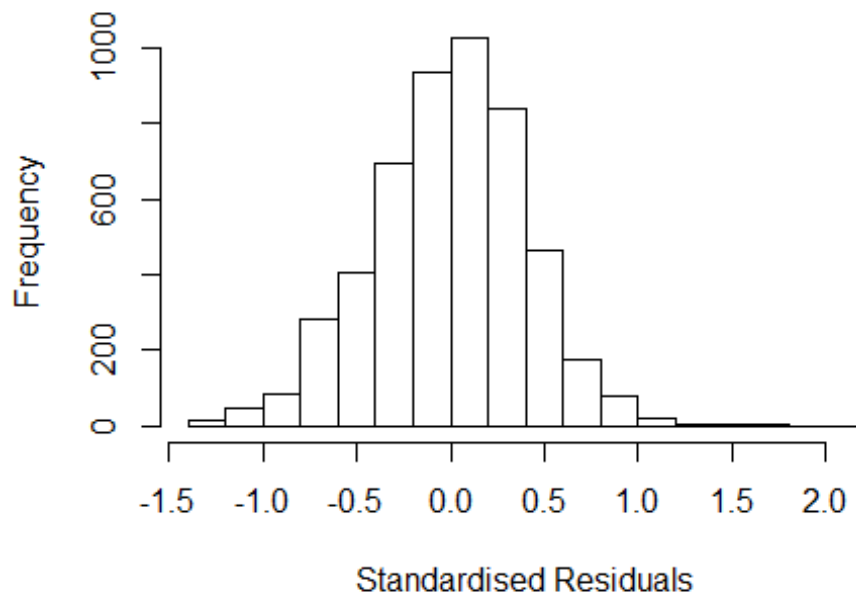


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 21, df = 1, p-value = 5e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 4.0548611199397"
## [1] "Male first author team size 2018 geometric mean: 3.60124368430352"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.26550691549458"
## [1] "Male last author team size 2018 geometric mean: 3.97749123224521"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8800, p-value = 0.003
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028 1          1.014
## LastAuthorFemale  1.013 1          1.006
## UniqueAuthors    1.134 4          1.016
## Year              1.151 16         1.004
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3414 -0.2632 0.0154 0.2644 2.1330
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.87720 0.03871 22.66 < 2e-16 ***
## FirstAuthorFemale1 -0.01725 0.01169 -1.48 0.14006
## LastAuthorFemale1 -0.01404 0.01384 -1.01 0.31030
## UniqueAuthors2 0.33833 0.03059 11.06 < 2e-16 ***
## UniqueAuthors3 0.41935 0.03020 13.89 < 2e-16 ***
## UniqueAuthors4 0.44315 0.03088 14.35 < 2e-16 ***
## UniqueAuthors5 0.53648 0.02972 18.05 < 2e-16 ***
## Year1997 -0.01819 0.04260 -0.43 0.66940
## Year1998 -0.00691 0.04401 -0.16 0.87523
## Year1999 -0.07225 0.04116 -1.76 0.07922 .
```

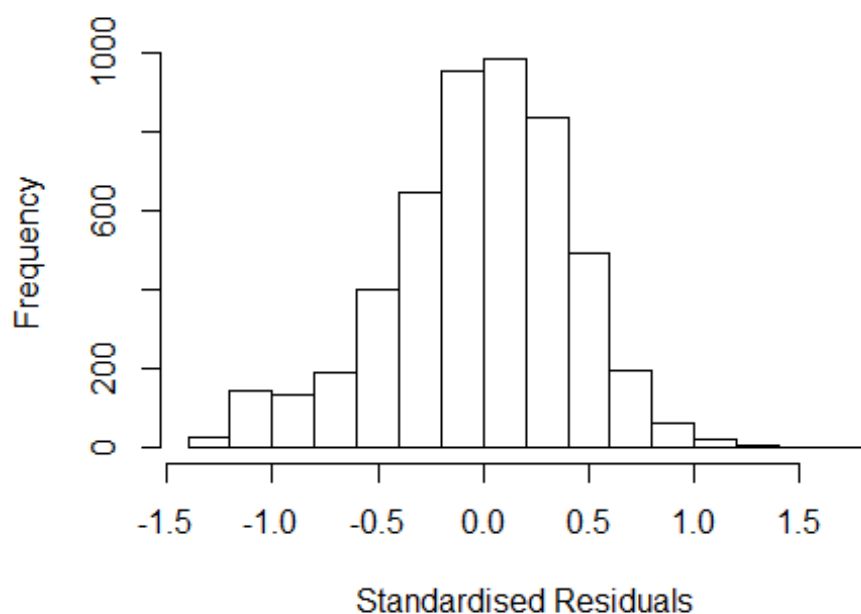


```

## Year2000      -0.09596      0.04144      -2.32      0.02061 *
## Year2001      -0.16365      0.04293      -3.81      0.00014 ***
## Year2002      -0.11320      0.03914      -2.89      0.00384 **
## Year2003      -0.14235      0.03914      -3.64      0.00028 ***
## Year2004      -0.12230      0.03936      -3.11      0.00190 **
## Year2005      -0.12510      0.03870      -3.23      0.00124 **
## Year2006      -0.09328      0.03698      -2.52      0.01170 *
## Year2007      -0.11305      0.03632      -3.11      0.00186 **
## Year2008      -0.11356      0.03777      -3.01      0.00265 **
## Year2009      -0.10067      0.03891      -2.59      0.00970 **
## Year2010      -0.09951      0.04016      -2.48      0.01326 *
## Year2011      -0.13276      0.03930      -3.38      0.00073 ***
## Year2012      -0.12158      0.03782      -3.21      0.00131 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.39
## Multiple R-squared:  0.134, Adjusted R-squared:  0.13
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 1661 is an outlier with |weight| = 0 ( < 2e-05);
## 449 weights are ~= 1. The remaining 4638 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0163 0.8640 0.9490 0.8980 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.97e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1 1.015
## LastAuthorFemale 1.015 1 1.007
## Year 1.043 16 1.001

```

## Residuals from first and last author

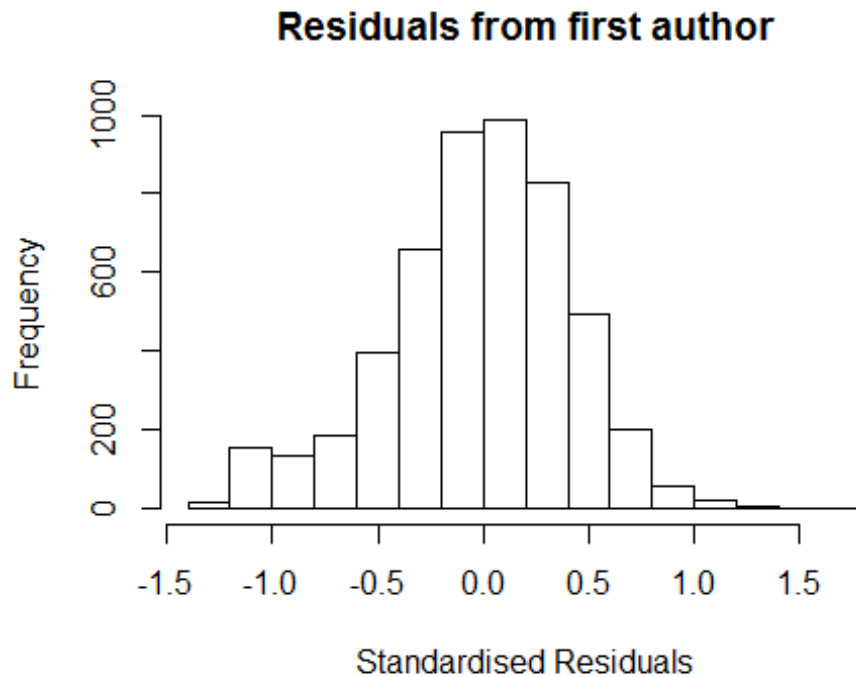


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2408 -0.2707  0.0107  0.2783  1.7493
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20271    0.03233   37.20  <2e-16 ***
## FirstAuthorFemale1  0.00992    0.01222    0.81    0.42
## LastAuthorFemale1 -0.02230    0.01454   -1.53    0.13
## Year1997         0.02816    0.04337    0.65    0.52
## Year1998         0.01472    0.04256    0.35    0.73
## Year1999        -0.01801    0.04200   -0.43    0.67
## Year2000        -0.03255    0.04162   -0.78    0.43
## Year2001        -0.06799    0.04584   -1.48    0.14
## Year2002        -0.05503    0.04087   -1.35    0.18
## Year2003        -0.06196    0.04020   -1.54    0.12
## Year2004        -0.05010    0.04059   -1.23    0.22
## Year2005        -0.04458    0.04025   -1.11    0.27
```

```

## Year2006      -0.02481    0.03790   -0.65    0.51
## Year2007      -0.03311    0.03686   -0.90    0.37
## Year2008      -0.02293    0.03809   -0.60    0.55
## Year2009      -0.01131    0.03936   -0.29    0.77
## Year2010      -0.01060    0.04072   -0.26    0.79
## Year2011      -0.05153    0.04085   -1.26    0.21
## Year2012      -0.03104    0.03879   -0.80    0.42
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.00388,    Adjusted R-squared:  0.000341
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 430 weights are ~= 1. The remaining 4658 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.021  0.866  0.949  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1      1.014
## Year              1.029 16      1.001

```

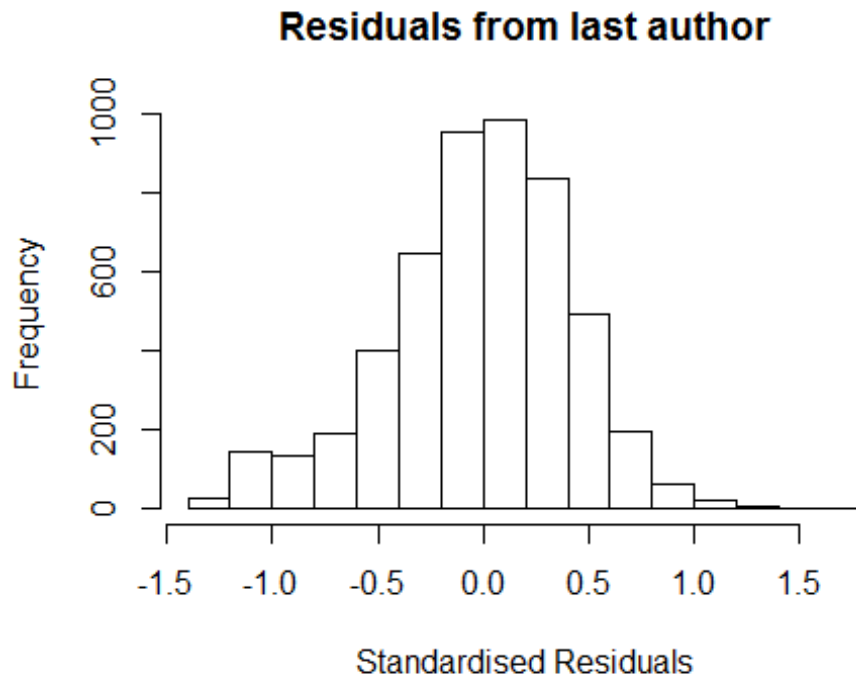


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.23609 -0.27246 0.00883 0.27430 1.75349
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.19906 0.03217 37.27 <2e-16 ***
## FirstAuthorFemale1 0.00842 0.01228 0.69 0.49
## Year1997 0.02861 0.04335 0.66 0.51
## Year1998 0.01397 0.04255 0.33 0.74
## Year1999 -0.01764 0.04201 -0.42 0.67
## Year2000 -0.03184 0.04166 -0.76 0.44
## Year2001 -0.06789 0.04588 -1.48 0.14
## Year2002 -0.05555 0.04084 -1.36 0.17
## Year2003 -0.06278 0.04016 -1.56 0.12
## Year2004 -0.04967 0.04060 -1.22 0.22
## Year2005 -0.04549 0.04028 -1.13 0.26
## Year2006 -0.02564 0.03790 -0.68 0.50
```

```

## Year2007          -0.03373    0.03687   -0.91    0.36
## Year2008          -0.02336    0.03812   -0.61    0.54
## Year2009          -0.01164    0.03938   -0.30    0.77
## Year2010          -0.01116    0.04069   -0.27    0.78
## Year2011          -0.05242    0.04085   -1.28    0.20
## Year2012          -0.03221    0.03883   -0.83    0.41
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.00346,    Adjusted R-squared:  0.000117
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 431 weights are ~= 1. The remaining 4657 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.020  0.867  0.950  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.014 1          1.007
## Year              1.014 16          1.000

```



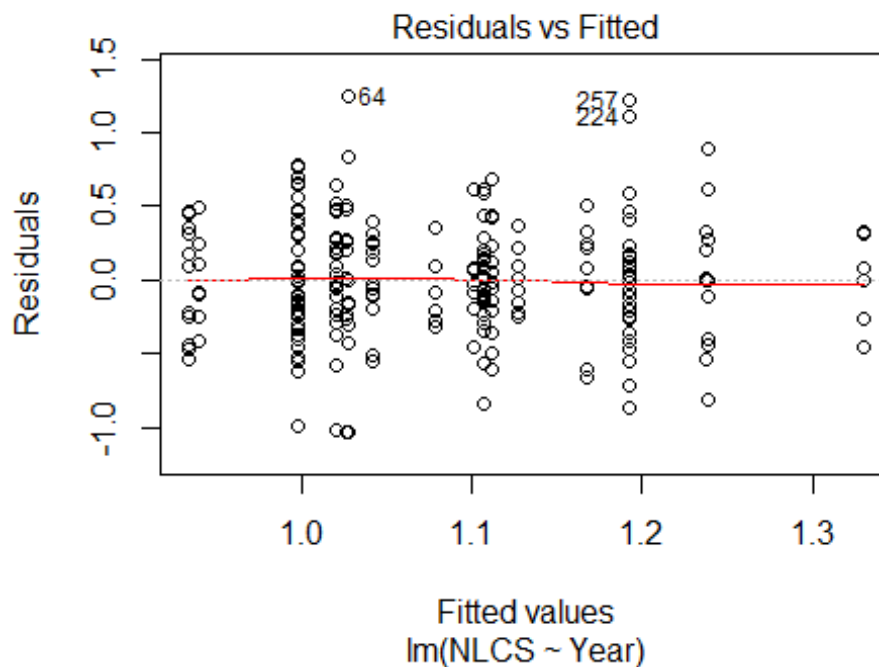
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2339 -0.2699  0.0103  0.2772  1.7454
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.20586    0.03193   37.77  <2e-16 ***
## LastAuthorFemale1 -0.02133    0.01461   -1.46    0.14
## Year1997         0.02806    0.04339    0.65    0.52
## Year1998         0.01382    0.04249    0.33    0.75
## Year1999        -0.01784    0.04198   -0.43    0.67
## Year2000        -0.03207    0.04163   -0.77    0.44
## Year2001        -0.06752    0.04581   -1.47    0.14
## Year2002        -0.05426    0.04084   -1.33    0.18
## Year2003        -0.06167    0.04019   -1.53    0.12
## Year2004        -0.04928    0.04061   -1.21    0.23
## Year2005        -0.04402    0.04025   -1.09    0.27
## Year2006        -0.02410    0.03786   -0.64    0.52
```

```

## Year2007          -0.03247      0.03686   -0.88      0.38
## Year2008          -0.02276      0.03809   -0.60      0.55
## Year2009          -0.01084      0.03934   -0.28      0.78
## Year2010          -0.00974      0.04071   -0.24      0.81
## Year2011          -0.05019      0.04080   -1.23      0.22
## Year2012          -0.03034      0.03879   -0.78      0.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.404
## Multiple R-squared:  0.00374,    Adjusted R-squared:  0.000399
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 432 weights are ~= 1. The remaining 4656 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.022  0.866  0.948  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.97e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5088"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2801"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   12   12   12   11   17   13   12   10   17   15   16   15   41   33
## 2011 2012
##   42   52
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8    9    8    5    8    6    8    7    7   13   15   11   12   26   23
## 2011 2012

```

```
## 32 40
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 8 8 5 8 6 8 6 6 8 13 10 12 24 22
## 2011 2012
## 30 39
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```

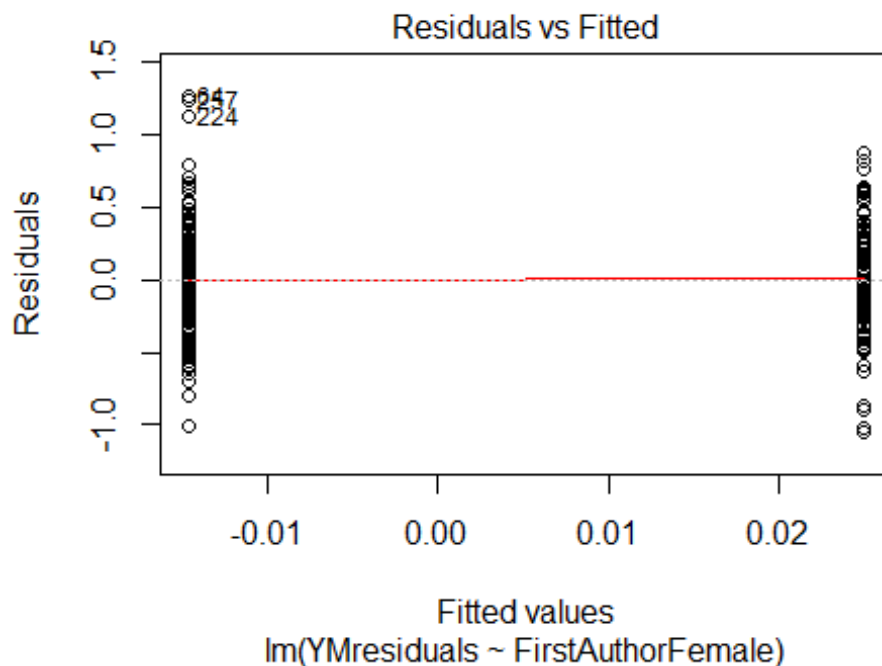


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.086, df = 1, p-value = 0.8
## [1] "Female first author team size 2018 geometric mean: 4.79238638400133"
## [1] "Male first author team size 2018 geometric mean: 4.00933320677449"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 130, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.6313753342337"
## [1] "Male last author team size 2018 geometric mean: 4.1933044975552"

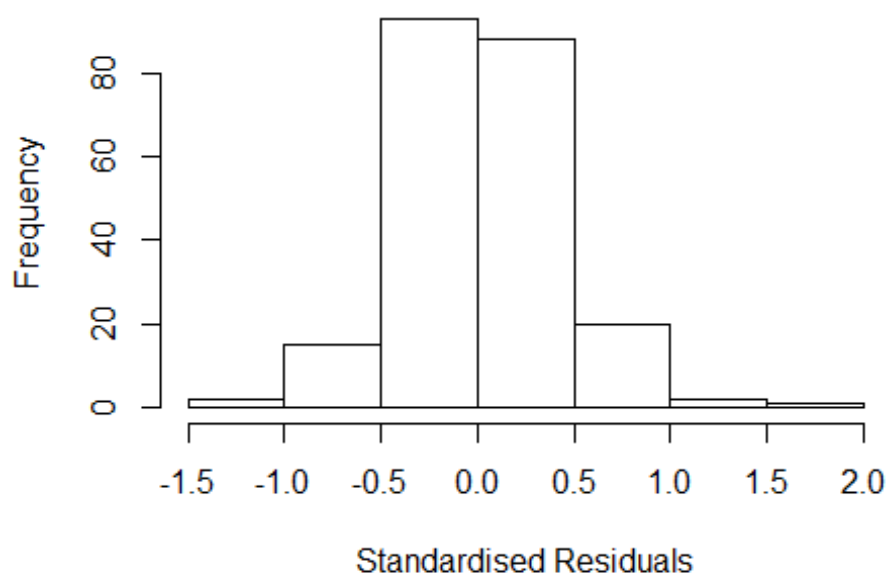
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.378	1	1.174
LastAuthorFemale	1.795	1	1.340
UniqueAuthors	14.296	4	1.394
Year	20.158	16	1.098

## Residuals from first and last author and team size



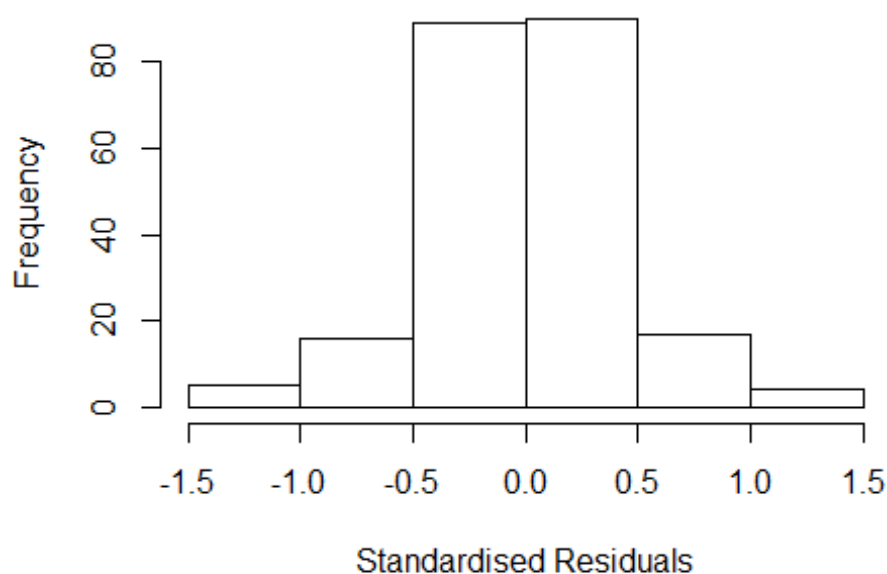
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.093837 -0.242086 0.000244 0.259708 1.509171
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1116 0.0648 17.14 <2e-16 ***
## FirstAuthorFemale1 0.0412 0.0590 0.70 0.486
## LastAuthorFemale1 0.1522 0.0706 2.16 0.032 *
## UniqueAuthors2 -0.0362 0.1162 -0.31 0.756
## UniqueAuthors3 0.0485 0.1214 0.40 0.690
## UniqueAuthors4 0.0545 0.1365 0.40 0.690
## UniqueAuthors5 0.3037 0.1223 2.48 0.014 *
## Year1997 -0.0216 0.1860 -0.12 0.908
## Year1998 0.0530 0.2289 0.23 0.817
## Year1999 0.0325 0.1900 0.17 0.864
```

```

## Year2000          -0.3269      0.2117    -1.54      0.124
## Year2001           0.1192      0.1429      0.83      0.405
## Year2002          -0.1485      0.1212    -1.22      0.222
## Year2003          -0.1810      0.1369    -1.32      0.188
## Year2004          -0.1287      0.2047    -0.63      0.530
## Year2005          -0.2533      0.1395    -1.82      0.071 .
## Year2006          -0.1786      0.1431    -1.25      0.214
## Year2007          -0.1610      0.1369    -1.18      0.241
## Year2008          -0.3071      0.1446    -2.12      0.035 *
## Year2009          -0.2098      0.1524    -1.38      0.170
## Year2010          -0.2028      0.1411    -1.44      0.152
## Year2011          -0.2658      0.1505    -1.77      0.079 .
## Year2012          -0.2656      0.1160    -2.29      0.023 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.371
## Multiple R-squared:  0.173, Adjusted R-squared:  0.081
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.060  0.891  0.950  0.898  0.989  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.347 1          1.160
## LastAuthorFemale  1.601 1          1.265
## Year              1.755 16          1.018

```

## Residuals from first and last author



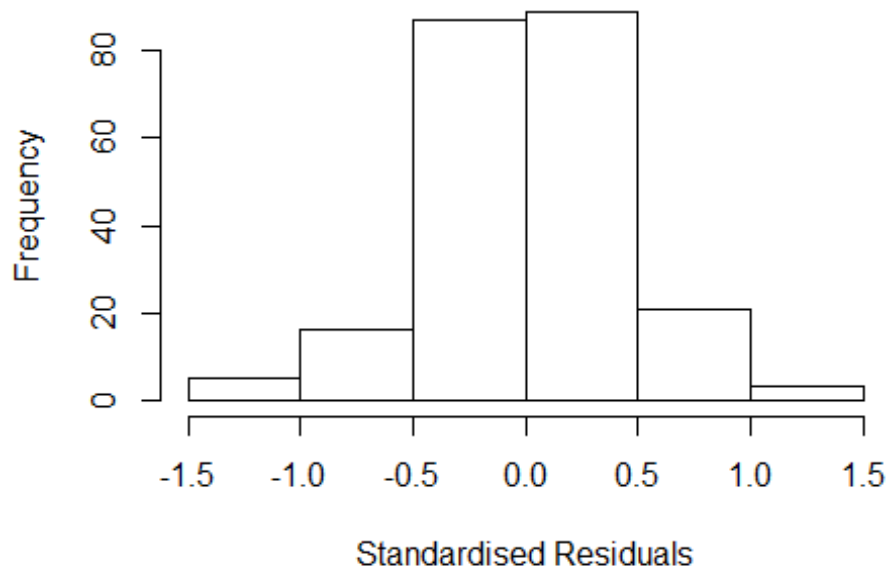
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.09411 -0.25840 0.00452 0.26011 1.49116
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10501 0.06149 17.97 <2e-16 ***
## FirstAuthorFemale1 0.02738 0.06198 0.44 0.659
## LastAuthorFemale1 0.15022 0.07079 2.12 0.035 *
## Year1997 -0.02145 0.16283 -0.13 0.895
## Year1998 0.10019 0.22372 0.45 0.655
## Year1999 0.12865 0.15121 0.85 0.396
## Year2000 -0.32317 0.23592 -1.37 0.172
## Year2001 0.22439 0.13812 1.62 0.106
## Year2002 -0.11677 0.11498 -1.02 0.311
## Year2003 -0.13966 0.11268 -1.24 0.217
## Year2004 -0.09875 0.16345 -0.60 0.546
## Year2005 -0.12551 0.15762 -0.80 0.427
```

```

## Year2006      -0.00798    0.11825   -0.07    0.946
## Year2007      -0.03865    0.14224   -0.27    0.786
## Year2008      -0.20535    0.13565   -1.51    0.132
## Year2009      -0.01046    0.11254   -0.09    0.926
## Year2010      -0.01559    0.09763   -0.16    0.873
## Year2011      -0.08752    0.10661   -0.82    0.413
## Year2012      -0.18850    0.09444   -2.00    0.047 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.402
## Multiple R-squared:  0.0848, Adjusted R-squared:  0.00322
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  0.883   0.951   0.903   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.219 1      1.104
## Year              1.219 16      1.006

```

## Residuals from first author



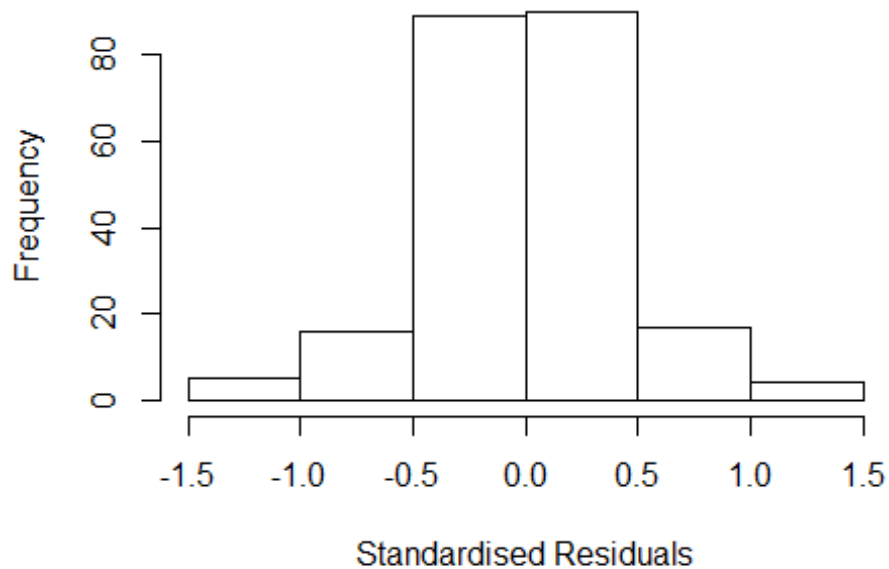
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1142 -0.2599 0.0218 0.2843 1.4267
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1172 0.0678 16.47 <2e-16 ***
## FirstAuthorFemale1 0.0740 0.0599 1.23 0.218
## Year1997 0.0268 0.1671 0.16 0.873
## Year1998 0.0797 0.2316 0.34 0.731
## Year1999 0.0861 0.1467 0.59 0.558
## Year2000 -0.2709 0.2394 -1.13 0.259
## Year2001 0.2031 0.1413 1.44 0.152
## Year2002 -0.0758 0.1139 -0.67 0.507
## Year2003 -0.1464 0.1198 -1.22 0.223
## Year2004 -0.0850 0.1640 -0.52 0.605
## Year2005 -0.1152 0.1578 -0.73 0.466
## Year2006 0.0188 0.1227 0.15 0.879
```

```

## Year2007          -0.0288      0.1386   -0.21    0.835
## Year2008          -0.1921      0.1309   -1.47    0.144
## Year2009           0.0154      0.1188    0.13    0.897
## Year2010          -0.0391      0.1018   -0.38    0.702
## Year2011          -0.0770      0.1165   -0.66    0.509
## Year2012          -0.1683      0.1000   -1.68    0.094 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.0609, Adjusted R-squared:  -0.0177
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 203 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.202  0.883   0.954   0.907   0.988   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.43 1      1.196
## Year              1.43 16      1.011

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.082488 -0.263629 -0.000545 0.274981 1.484189
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10709 0.06234 17.76 <2e-16 ***
## LastAuthorFemale1 0.16041 0.06682 2.40 0.017 *
## Year1997 -0.02398 0.16215 -0.15 0.883
## Year1998 0.10479 0.22676 0.46 0.644
## Year1999 0.14414 0.14879 0.97 0.334
## Year2000 -0.31828 0.23876 -1.33 0.184
## Year2001 0.22754 0.13764 1.65 0.100 .
## Year2002 -0.11128 0.11714 -0.95 0.343
## Year2003 -0.13087 0.11447 -1.14 0.254
## Year2004 -0.08955 0.15917 -0.56 0.574
## Year2005 -0.11838 0.15452 -0.77 0.445
## Year2006 -0.00549 0.11852 -0.05 0.963
```

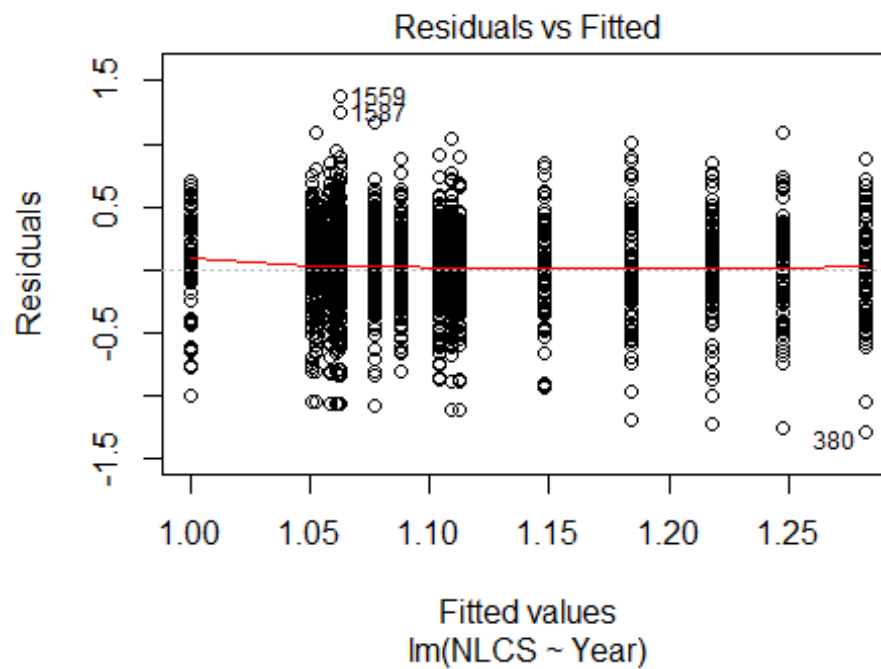


```

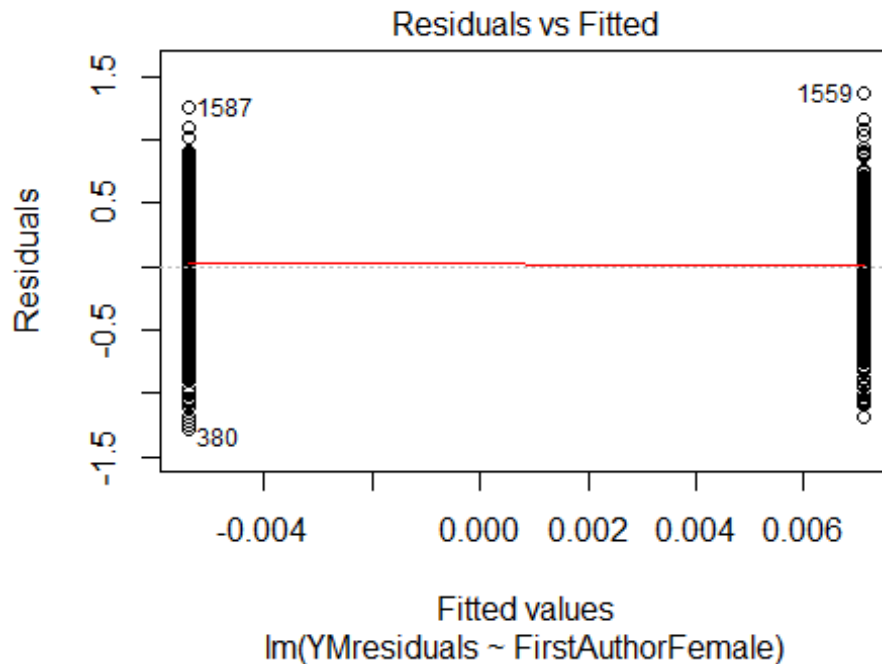
## Year2007          -0.03698      0.14400    -0.26      0.798
## Year2008          -0.20414      0.13718    -1.49      0.138
## Year2009          -0.00558      0.11268    -0.05      0.961
## Year2010          -0.00376      0.09600    -0.04      0.969
## Year2011          -0.08433      0.10674    -0.79      0.430
## Year2012          -0.18502      0.09559    -1.94      0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.0837, Adjusted R-squared:  0.00695
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.145  0.881  0.952  0.903  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 221"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2802"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 156 159 138 118 152 154 186 152 161 203 221 241 210 213 224
## 2011 2012
## 222 217
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 95 91 80 61 81 71 131 105 99 141 165 181 154 165 168
## 2011 2012

```

```
## 165 158
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 85 80 73 56 75 62 110 97 86 127 144 165 135 147 147
## 2011 2012
## 153 144
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 79, df = 16, p-value = 2e-10
```

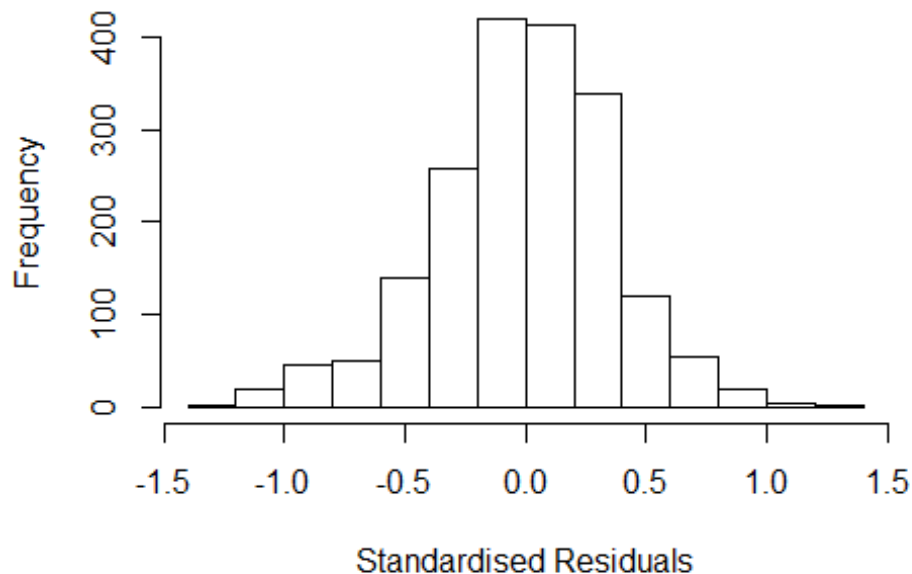


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 18, df = 1, p-value = 3e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 3.41874794179304"
## [1] "Male first author team size 2018 geometric mean: 3.32922557654426"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.60330303030796"
## [1] "Male last author team size 2018 geometric mean: 3.21685869353691"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2900, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.080 1          1.039
## LastAuthorFemale  1.104 1          1.051
## UniqueAuthors    1.258 4          1.029
## Year             1.439 16          1.011
```

## Residuals from first and last author and team size



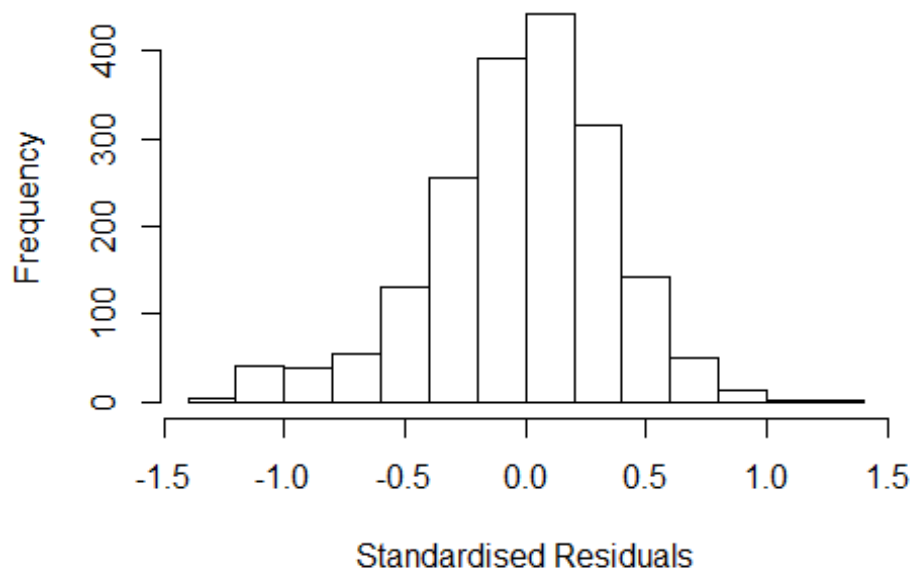
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33137 -0.22972  0.00386  0.23285  1.39576
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0783     0.0454   23.72 < 2e-16 ***
## FirstAuthorFemale1  0.0064     0.0167    0.38  0.70220
## LastAuthorFemale1 -0.0226     0.0184   -1.23  0.21945
## UniqueAuthors2     0.1749     0.0354    4.94  8.4e-07 ***
## UniqueAuthors3     0.2531     0.0350    7.24  6.6e-13 ***
## UniqueAuthors4     0.2773     0.0365    7.59  4.9e-14 ***
## UniqueAuthors5     0.2892     0.0358    8.08  1.1e-15 ***
## Year1997          -0.0199     0.0639   -0.31  0.75536
## Year1998           0.0149     0.0652    0.23  0.81949
## Year1999          -0.0964     0.0660   -1.46  0.14468
```

```

## Year2000          -0.0609      0.0645   -0.94   0.34493
## Year2001          -0.1668      0.0701   -2.38   0.01735 *
## Year2002          -0.1828      0.0539   -3.39   0.00071 ***
## Year2003          -0.1705      0.0617   -2.77   0.00574 **
## Year2004          -0.2174      0.0601   -3.62   0.00031 ***
## Year2005          -0.1955      0.0490   -3.99   6.9e-05 ***
## Year2006          -0.1791      0.0460   -3.89   0.00010 ***
## Year2007          -0.1747      0.0435   -4.02   6.1e-05 ***
## Year2008          -0.1806      0.0451   -4.00   6.6e-05 ***
## Year2009          -0.2304      0.0466   -4.94   8.5e-07 ***
## Year2010          -0.2395      0.0477   -5.02   5.7e-07 ***
## Year2011          -0.2030      0.0489   -4.15   3.5e-05 ***
## Year2012          -0.2183      0.0477   -4.58   5.0e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.34
## Multiple R-squared:  0.0987, Adjusted R-squared:  0.0881
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 182 weights are ~= 1. The remaining 1704 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0531 0.8660 0.9480 0.8920 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.30e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.067 1 1.033
## LastAuthorFemale 1.086 1 1.042
## Year 1.152 16 1.004

```

## Residuals from first and last author



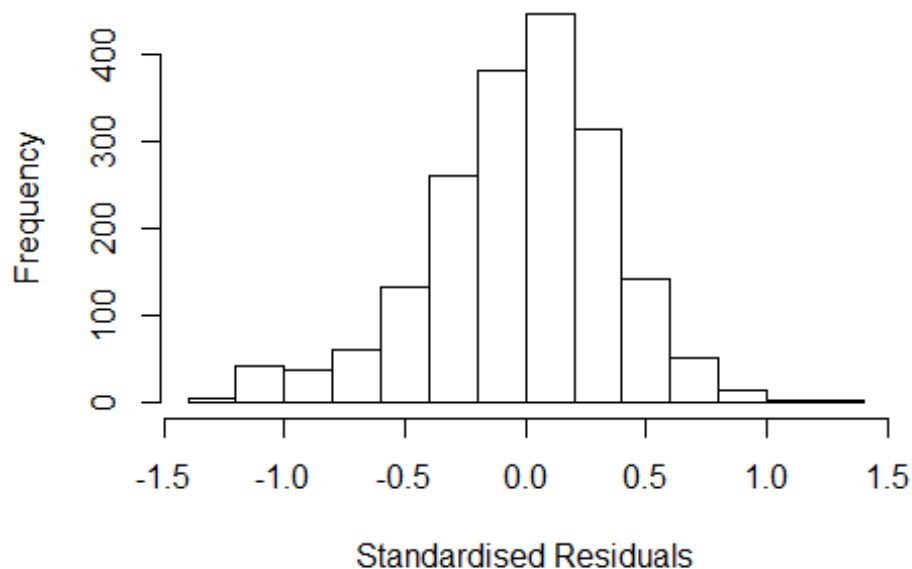
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2861 -0.2343  0.0153  0.2327  1.3516
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2429    0.0391   31.77 < 2e-16 ***
## FirstAuthorFemale1  0.0214    0.0171    1.25  0.2112
## LastAuthorFemale1 -0.0326    0.0186   -1.75  0.0803 .
## Year1997          0.0069    0.0628    0.11  0.9125
## Year1998          0.0432    0.0661    0.65  0.5136
## Year1999         -0.0730    0.0675   -1.08  0.2802
## Year2000         -0.0338    0.0676   -0.50  0.6172
## Year2001         -0.1278    0.0701   -1.83  0.0681 .
## Year2002         -0.1417    0.0549   -2.58  0.0100 *
## Year2003         -0.1338    0.0617   -2.17  0.0302 *
## Year2004         -0.1780    0.0597   -2.98  0.0029 **
## Year2005         -0.1498    0.0498   -3.01  0.0027 **
```

```

## Year2006          -0.1364      0.0468   -2.91   0.0036 **
## Year2007          -0.1340      0.0451   -2.97   0.0030 **
## Year2008          -0.1235      0.0452   -2.73   0.0063 **
## Year2009          -0.1840      0.0470   -3.92   9.3e-05 ***
## Year2010          -0.1867      0.0473   -3.95   8.1e-05 ***
## Year2011          -0.1539      0.0487   -3.16   0.0016 **
## Year2012          -0.1504      0.0476   -3.16   0.0016 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.0214
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 158 weights are ~= 1. The remaining 1728 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.101  0.868  0.949  0.893  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1      1.032
## Year              1.065 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2818 -0.2357 0.0103 0.2355 1.3618
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23370 0.03849 32.05 < 2e-16 ***
## FirstAuthorFemale1 0.01858 0.01736 1.07 0.2846
## Year1997 0.00944 0.06288 0.15 0.8806
## Year1998 0.04811 0.06588 0.73 0.4654
## Year1999 -0.06584 0.06659 -0.99 0.3229
## Year2000 -0.03280 0.06780 -0.48 0.6286
## Year2001 -0.12305 0.06990 -1.76 0.0785 .
## Year2002 -0.14003 0.05483 -2.55 0.0107 *
## Year2003 -0.13253 0.06205 -2.14 0.0328 *
## Year2004 -0.17605 0.05975 -2.95 0.0033 **
## Year2005 -0.15044 0.04999 -3.01 0.0027 **
## Year2006 -0.13463 0.04680 -2.88 0.0041 **
```

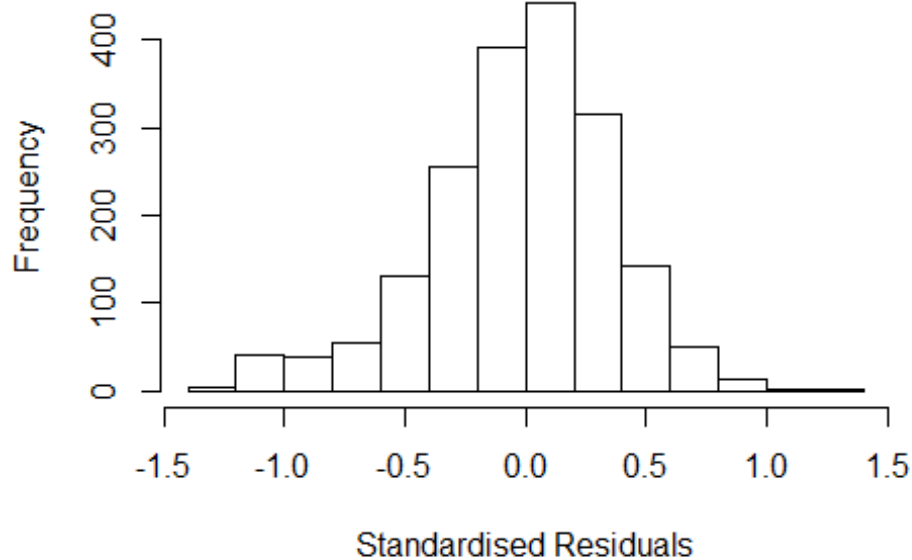


```

## Year2007          -0.13589    0.04515   -3.01    0.0027 **
## Year2008          -0.12329    0.04512   -2.73    0.0063 **
## Year2009          -0.18446    0.04708   -3.92    9.3e-05 ***
## Year2010          -0.18827    0.04738   -3.97    7.4e-05 ***
## Year2011          -0.15467    0.04887   -3.16    0.0016 **
## Year2012          -0.15180    0.04752   -3.19    0.0014 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.349
## Multiple R-squared:  0.0293, Adjusted R-squared:  0.0205
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 165 weights are ~= 1. The remaining 1721 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0943 0.8680 0.9490 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.083 1          1.041
## Year              1.083 16          1.002

```

## Residuals from last author



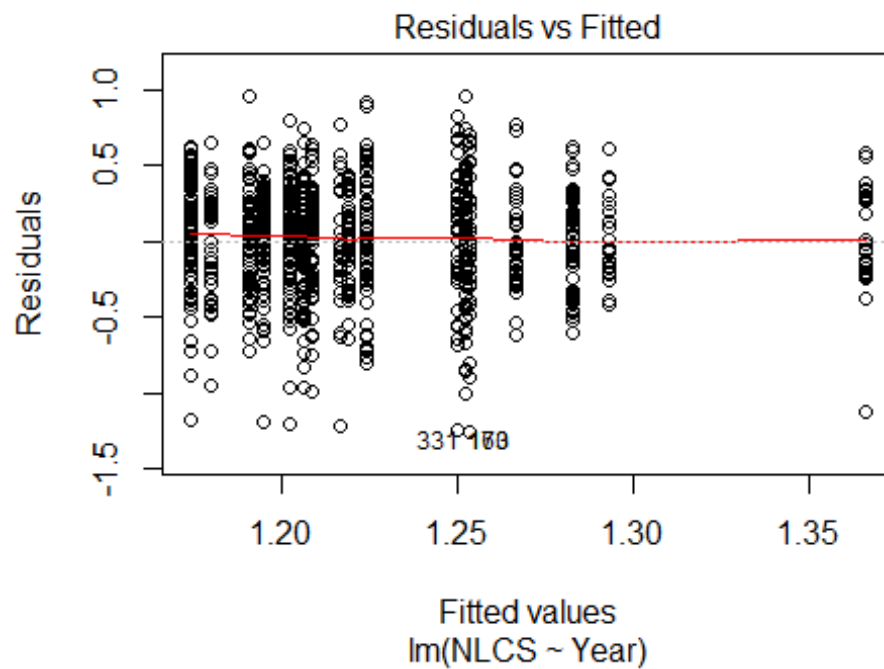
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.296 -0.236 0.015 0.229 1.362
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2504 0.0384 32.53 < 2e-16 ***
## LastAuthorFemale1 -0.0304 0.0188 -1.62 0.10590
## Year1997 0.0071 0.0629 0.11 0.91009
## Year1998 0.0452 0.0659 0.68 0.49352
## Year1999 -0.0727 0.0673 -1.08 0.28007
## Year2000 -0.0365 0.0677 -0.54 0.58976
## Year2001 -0.1288 0.0701 -1.84 0.06626 .
## Year2002 -0.1379 0.0546 -2.53 0.01160 *
## Year2003 -0.1332 0.0616 -2.16 0.03056 *
## Year2004 -0.1748 0.0595 -2.94 0.00336 **
## Year2005 -0.1497 0.0498 -3.00 0.00269 **
## Year2006 -0.1336 0.0467 -2.86 0.00427 **
```

```

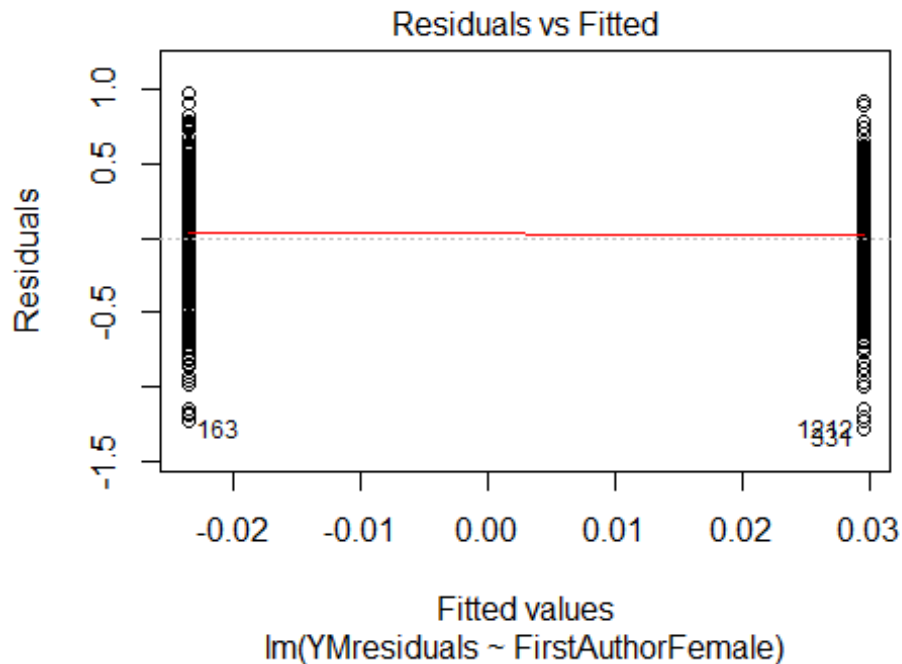
## Year2007          -0.1340      0.0450   -2.98  0.00293 **
## Year2008          -0.1226      0.0452   -2.71  0.00681 **
## Year2009          -0.1804      0.0468   -3.86  0.00012 ***
## Year2010          -0.1840      0.0471   -3.90  9.8e-05 ***
## Year2011          -0.1522      0.0486   -3.13  0.00176 **
## Year2012          -0.1485      0.0474   -3.14  0.00174 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.35
## Multiple R-squared:  0.03,   Adjusted R-squared:  0.0211
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 160 weights are ~= 1. The remaining 1726 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0949 0.8710 0.9480 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.30e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1886"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2803"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   69   77   60   48   64   61   60   67   68   77   74  107  108   95   83
## 2011 2012
##  104  112
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   25   41   27   22   35   38   37   42   50   62   62   85   82   82   66
## 2011 2012

```

```
##      84  101
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   38   25   20   28   32   29   39   44   55   54   73   71   75   52
## 2011 2012
##   73   93
## [1] "Heteroscedasticity checks, confirming that there are problems with
##      these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 33, df = 16, p-value = 0.007
```

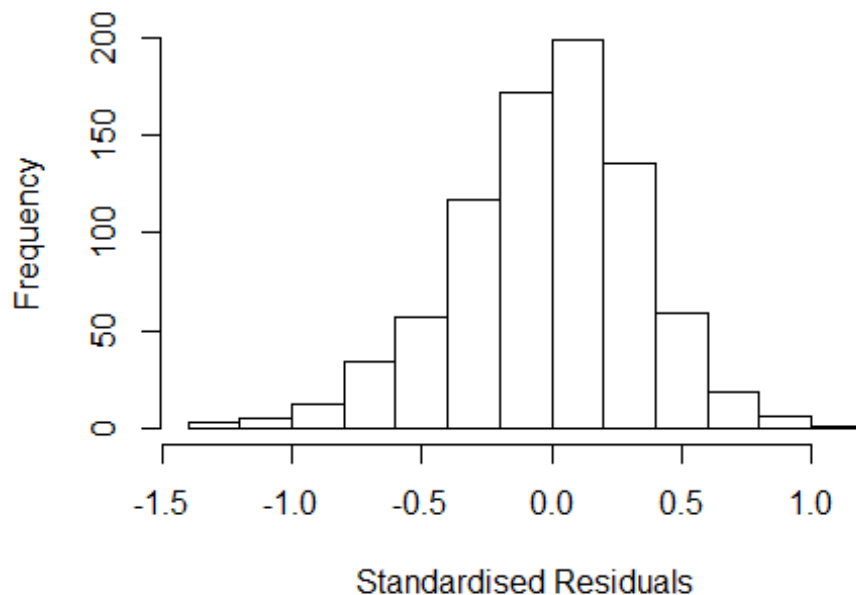


```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.4, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 4.9324923810703"
## [1] "Male first author team size 2018 geometric mean: 4.57211842153384"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.79294277963444"
## [1] "Male last author team size 2018 geometric mean: 4.76783859610134"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1700, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.158 1          1.076
## LastAuthorFemale  1.081 1          1.040
## UniqueAuthors     1.440 4          1.047
## Year               1.734 16         1.017
```

## Residuals from first and last author and team size



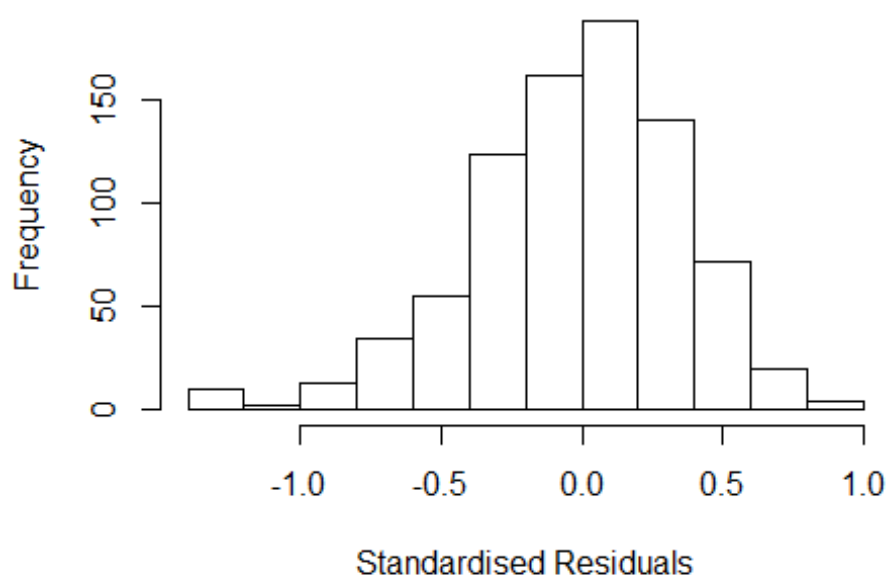
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30232 -0.22394 0.00975 0.21739 1.05946
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0476 0.1135 9.23 < 2e-16 ***
## FirstAuthorFemale1 0.0354 0.0257 1.38 0.169
## LastAuthorFemale1 0.0474 0.0265 1.79 0.074 .
## UniqueAuthors2 0.1493 0.0724 2.06 0.039 *
## UniqueAuthors3 0.1728 0.0686 2.52 0.012 *
## UniqueAuthors4 0.3028 0.0666 4.55 6.3e-06 ***
## UniqueAuthors5 0.2772 0.0629 4.41 1.2e-05 ***
## Year1997 0.0147 0.1487 0.10 0.921
## Year1998 0.1377 0.1352 1.02 0.308
## Year1999 0.0132 0.1359 0.10 0.923
```

```

## Year2000          -0.0193      0.1447   -0.13    0.894
## Year2001          -0.0529      0.1254   -0.42    0.673
## Year2002          -0.1171      0.1292   -0.91    0.365
## Year2003          -0.1143      0.1211   -0.94    0.345
## Year2004          -0.0451      0.1224   -0.37    0.713
## Year2005          -0.0989      0.1216   -0.81    0.416
## Year2006          -0.1231      0.1227   -1.00    0.316
## Year2007          -0.1044      0.1207   -0.87    0.387
## Year2008          -0.0491      0.1225   -0.40    0.689
## Year2009          -0.0524      0.1222   -0.43    0.668
## Year2010          -0.0737      0.1313   -0.56    0.575
## Year2011          -0.0570      0.1208   -0.47    0.637
## Year2012          -0.0944      0.1216   -0.78    0.438
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.328
## Multiple R-squared:  0.0813, Adjusted R-squared:  0.0559
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0802 0.8640 0.9480 0.8910 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.22e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.136 1 1.066
## LastAuthorFemale 1.055 1 1.027
## Year 1.195 16 1.006

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3192 -0.2408  0.0104  0.2322  0.9610
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19531    0.10950   10.92  <2e-16 ***
## FirstAuthorFemale1  0.04347    0.02645    1.64   0.101
## LastAuthorFemale1  0.04591    0.02746    1.67   0.095 .
## Year1997         0.07783    0.13438    0.58   0.563
## Year1998         0.17829    0.12765    1.40   0.163
## Year1999         0.04498    0.12725    0.35   0.724
## Year2000         0.06115    0.13868    0.44   0.659
## Year2001         0.01469    0.12379    0.12   0.906
## Year2002        -0.07633    0.13350   -0.57   0.568
## Year2003        -0.02864    0.11871   -0.24   0.809
## Year2004         0.04637    0.12198    0.38   0.704
## Year2005        -0.00348    0.11929   -0.03   0.977
```

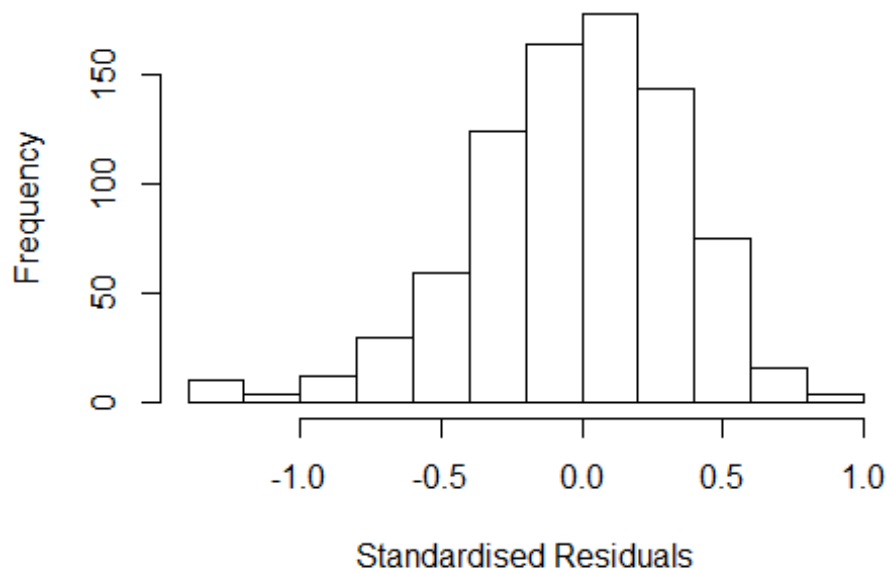


```

## Year2006      -0.04335    0.12109   -0.36    0.720
## Year2007      -0.02562    0.11828   -0.22    0.829
## Year2008       0.03454    0.11979    0.29    0.773
## Year2009       0.04768    0.12025    0.40    0.692
## Year2010       0.00941    0.13008    0.07    0.942
## Year2011       0.02788    0.11698    0.24    0.812
## Year2012      -0.03013    0.11930   -0.25    0.801
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.344
## Multiple R-squared:  0.0251, Adjusted R-squared:  0.0032
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 744 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.109  0.866  0.947  0.894  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.22e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.136 1         1.066
## Year              1.136 16         1.004

```

## Residuals from first author



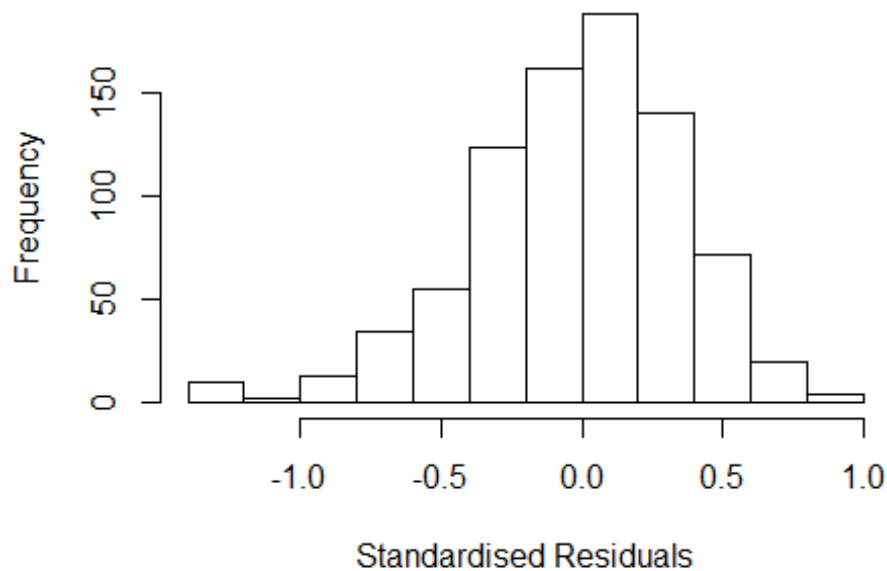
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28977 -0.24553 0.00583 0.23337 0.94367
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20e+00 1.09e-01 11.07 <2e-16 ***
## FirstAuthorFemale1 4.65e-02 2.65e-02 1.75 0.08 .
## Year1997 8.73e-02 1.34e-01 0.65 0.52
## Year1998 1.83e-01 1.26e-01 1.44 0.15
## Year1999 5.01e-02 1.26e-01 0.40 0.69
## Year2000 6.77e-02 1.37e-01 0.50 0.62
## Year2001 2.28e-02 1.22e-01 0.19 0.85
## Year2002 -7.15e-02 1.31e-01 -0.54 0.59
## Year2003 -1.25e-02 1.17e-01 -0.11 0.91
## Year2004 5.45e-02 1.21e-01 0.45 0.65
## Year2005 6.24e-05 1.18e-01 0.00 1.00
## Year2006 -3.59e-02 1.20e-01 -0.30 0.76
```

```

## Year2007          -1.76e-02   1.16e-01   -0.15    0.88
## Year2008          4.25e-02   1.18e-01    0.36    0.72
## Year2009          5.95e-02   1.18e-01    0.50    0.61
## Year2010          1.86e-02   1.29e-01    0.14    0.89
## Year2011          3.33e-02   1.15e-01    0.29    0.77
## Year2012          -2.07e-02   1.17e-01   -0.18    0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.346
## Multiple R-squared:  0.0212, Adjusted R-squared:  0.000503
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 743 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.133  0.865  0.947  0.895  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.22e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.047 1          1.023
## Year            1.047 16          1.001

```

## Residuals from last author



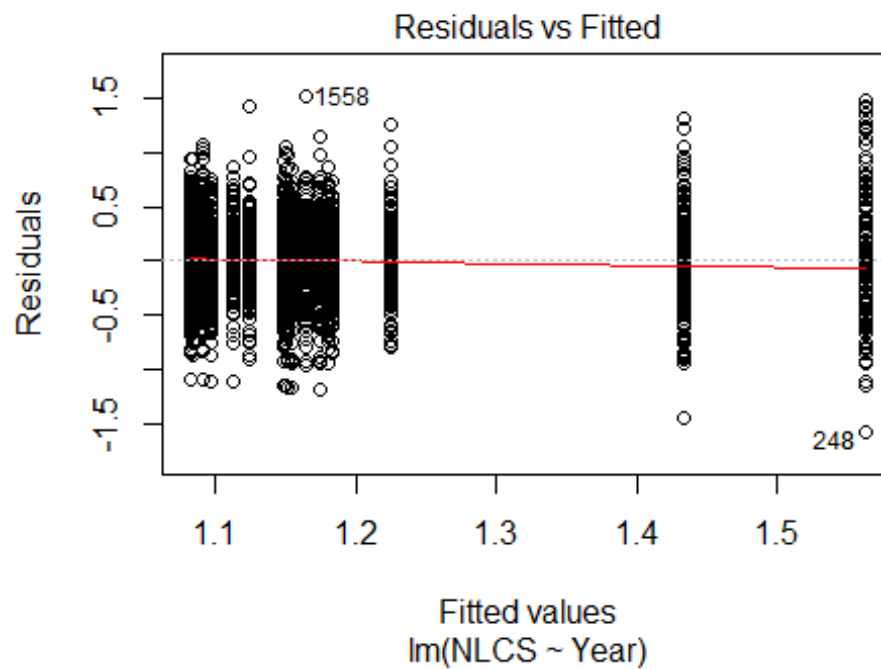
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.341 -0.245 0.015 0.232 0.945
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.20628 0.10962 11.00 <2e-16 ***
## LastAuthorFemale1 0.04889 0.02735 1.79 0.074 .
## Year1997 0.08534 0.13378 0.64 0.524
## Year1998 0.17724 0.12749 1.39 0.165
## Year1999 0.05106 0.12648 0.40 0.687
## Year2000 0.06822 0.13896 0.49 0.624
## Year2001 0.02203 0.12354 0.18 0.859
## Year2002 -0.06462 0.13321 -0.49 0.628
## Year2003 -0.02523 0.11832 -0.21 0.831
## Year2004 0.04979 0.12213 0.41 0.684
## Year2005 0.00637 0.11832 0.05 0.957
## Year2006 -0.03334 0.12001 -0.28 0.781
```

```

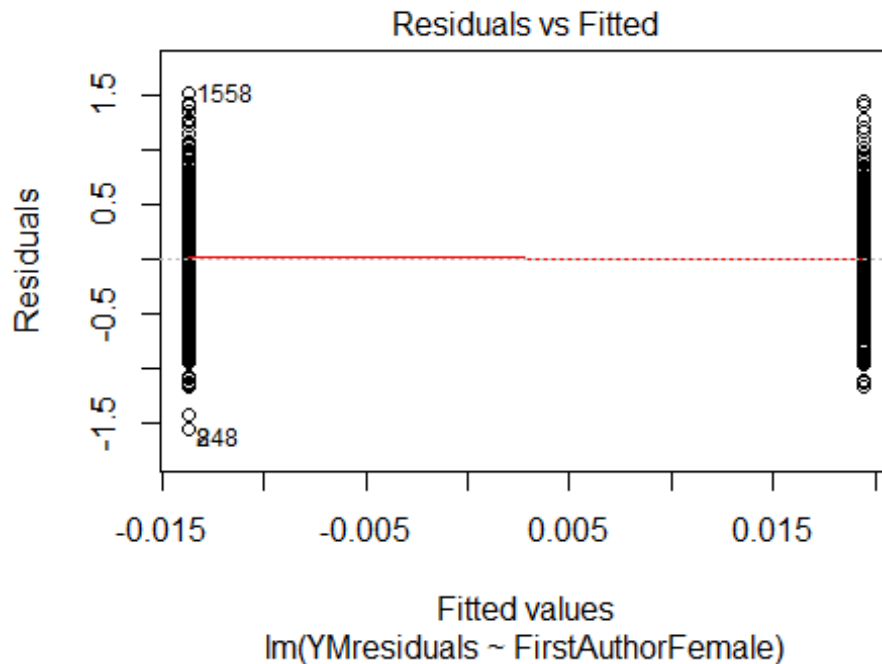
## Year2007          -0.02160      0.11790    -0.18      0.855
## Year2008           0.04559      0.11881      0.38      0.701
## Year2009           0.05255      0.11996      0.44      0.661
## Year2010           0.01773      0.12955      0.14      0.891
## Year2011           0.04368      0.11527      0.38      0.705
## Year2012          -0.01988      0.11851     -0.17      0.867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.343
## Multiple R-squared:  0.0215, Adjusted R-squared:  0.000741
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 750 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0934 0.8630 0.9500 0.8950 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.22e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 820"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2804"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 411 376 303 259 330 362 298 242 271 262 262 276 295 309 329
## 2011 2012
## 376 365
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 134 145 161 135 133 142 164 119 145 144 179 182 190 200 224
## 2011 2012

```

```
## 271 263
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 119 123 145 120 113 119 125 99 126 123 137 158 155 179 191
## 2011 2012
## 230 223
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```

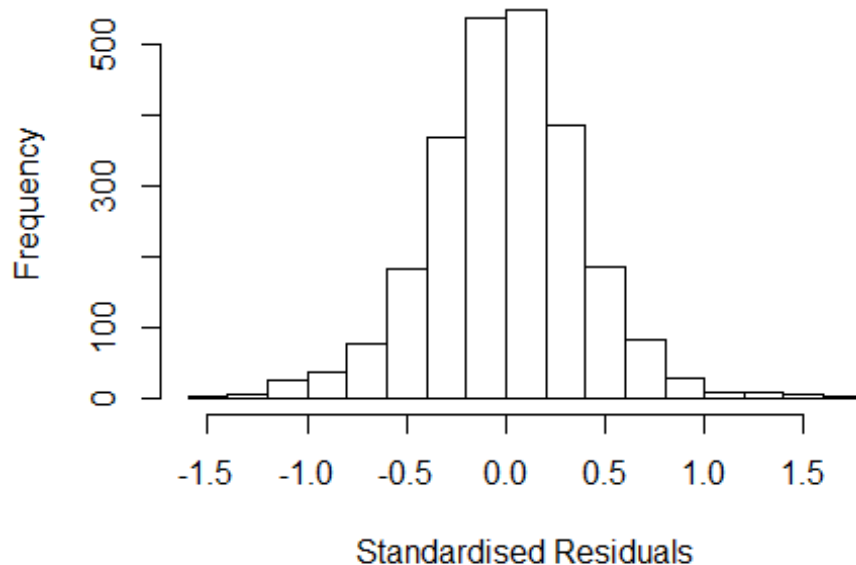


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 40, df = 1, p-value = 2e-10
```



```
## [1] "Female first author team size 2018 geometric mean: 5.39345808945417"
## [1] "Male first author team size 2018 geometric mean: 4.08557322497152"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6200, p-value = 0.002
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.28107951521479"
## [1] "Male last author team size 2018 geometric mean: 4.32143165562239"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5400, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.048 1      1.024
## LastAuthorFemale  1.056 1      1.028
## UniqueAuthors     1.149 4      1.017
## Year              1.192 16     1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.55347 -0.23854  0.00291  0.23911  1.70335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.2978    0.0740   17.55 < 2e-16 ***
## FirstAuthorFemale1  0.0107    0.0150    0.71  0.47668
## LastAuthorFemale1  0.0222    0.0179    1.24  0.21448
## UniqueAuthors2     0.1188    0.0446    2.67  0.00774 **
## UniqueAuthors3     0.1546    0.0436    3.55  0.00040 ***
## UniqueAuthors4     0.1688    0.0436    3.87  0.00011 ***
## UniqueAuthors5     0.2557    0.0417    6.13  1.0e-09 ***
## Year1997          -0.0126    0.0750   -0.17  0.86717
## Year1998          -0.2721    0.0679   -4.01  6.3e-05 ***
## Year1999          -0.3209    0.0684   -4.69  2.9e-06 ***
```

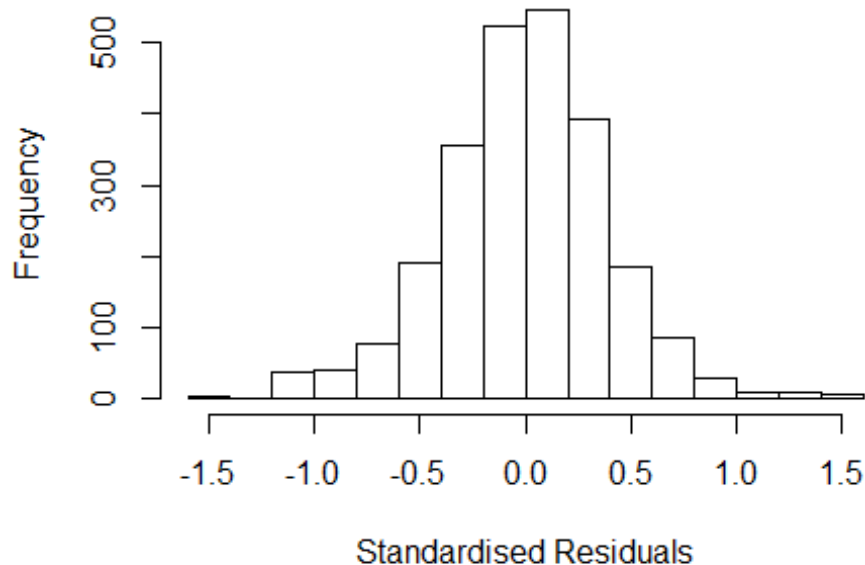


```

## Year2000      -0.2691      0.0686      -3.92      9.0e-05 ***
## Year2001      -0.3697      0.0672      -5.50      4.1e-08 ***
## Year2002      -0.3120      0.0689      -4.53      6.3e-06 ***
## Year2003      -0.3225      0.0686      -4.70      2.7e-06 ***
## Year2004      -0.3771      0.0679      -5.55      3.1e-08 ***
## Year2005      -0.4175      0.0689      -6.06      1.6e-09 ***
## Year2006      -0.3388      0.0685      -4.95      8.0e-07 ***
## Year2007      -0.3597      0.0666      -5.40      7.4e-08 ***
## Year2008      -0.3354      0.0665      -5.04      4.9e-07 ***
## Year2009      -0.3342      0.0669      -5.00      6.2e-07 ***
## Year2010      -0.3056      0.0659      -4.64      3.7e-06 ***
## Year2011      -0.3916      0.0662      -5.91      3.9e-09 ***
## Year2012      -0.3908      0.0668      -5.85      5.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.351
## Multiple R-squared:  0.102, Adjusted R-squared:  0.0943
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(112,114) are outliers with |weight| = 0 ( < 4e-05);
## 218 weights are ~= 1. The remaining 2265 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.006  0.869  0.951  0.892  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           4.02e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
##   nResample   max.it   best.r.s   k.fast.s   k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev   mts   compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1 1.021
## LastAuthorFemale 1.035 1 1.017
## Year 1.054 16 1.002

```

## Residuals from first and last author



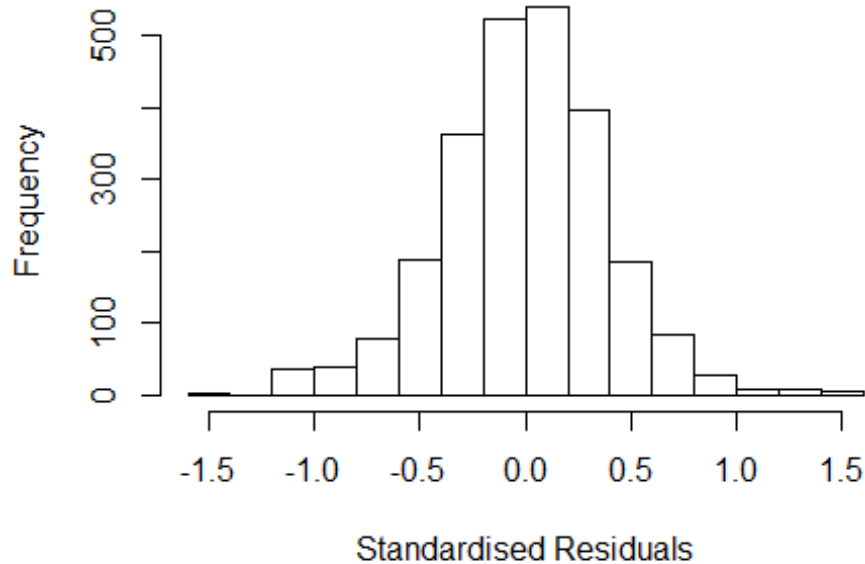
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.48360 -0.24367 0.00533 0.23962 1.53462
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4836 0.0642 23.12 < 2e-16 ***
## FirstAuthorFemale1 0.0222 0.0153 1.45 0.14758
## LastAuthorFemale1 0.0179 0.0180 1.00 0.31884
## Year1997 -0.0438 0.0770 -0.57 0.56974
## Year1998 -0.2956 0.0700 -4.22 2.5e-05 ***
## Year1999 -0.3432 0.0713 -4.81 1.6e-06 ***
## Year2000 -0.2718 0.0719 -3.78 0.00016 ***
## Year2001 -0.3786 0.0697 -5.43 6.1e-08 ***
## Year2002 -0.3149 0.0725 -4.34 1.5e-05 ***
## Year2003 -0.3265 0.0723 -4.51 6.6e-06 ***
## Year2004 -0.3925 0.0713 -5.51 4.0e-08 ***
## Year2005 -0.4181 0.0716 -5.84 5.9e-09 ***
```

```

## Year2006          -0.3367      0.0710    -4.74  2.2e-06 ***
## Year2007          -0.3675      0.0697    -5.27  1.5e-07 ***
## Year2008          -0.3407      0.0693    -4.92  9.4e-07 ***
## Year2009          -0.3267      0.0699    -4.68  3.1e-06 ***
## Year2010          -0.3032      0.0686    -4.42  1.0e-05 ***
## Year2011          -0.3909      0.0692    -5.65  1.8e-08 ***
## Year2012          -0.3977      0.0702    -5.67  1.6e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0703, Adjusted R-squared:  0.0635
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 217 weights are ~= 1. The remaining 2268 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0273 0.8680 0.9500 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.02e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.029 1      1.015
## Year      1.029 16      1.001

```

## Residuals from first author



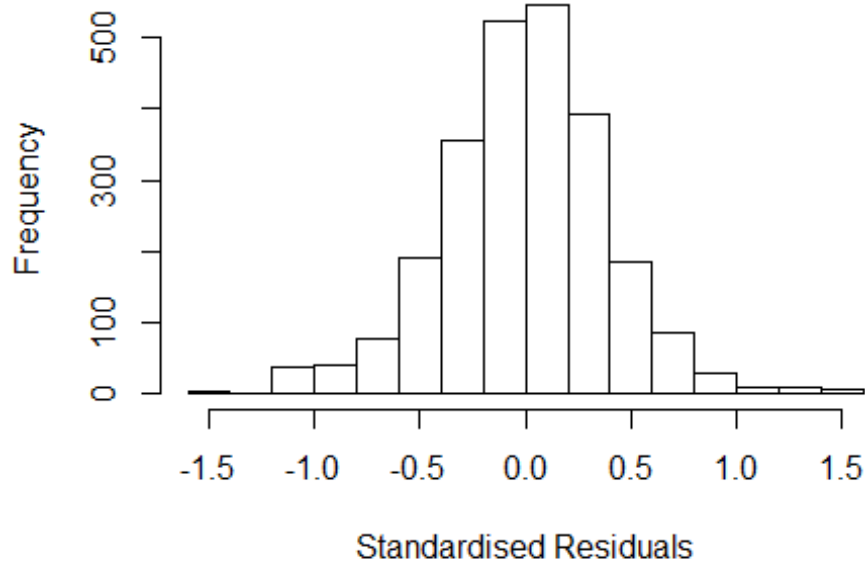
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.48694 -0.24441 0.00269 0.23959 1.53246
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4869 0.0637 23.34 < 2e-16 ***
## FirstAuthorFemale1 0.0239 0.0152 1.57 0.11746
## Year1997 -0.0445 0.0767 -0.58 0.56186
## Year1998 -0.2964 0.0698 -4.24 2.3e-05 ***
## Year1999 -0.3444 0.0711 -4.85 1.3e-06 ***
## Year2000 -0.2716 0.0718 -3.79 0.00016 ***
## Year2001 -0.3785 0.0695 -5.45 5.6e-08 ***
## Year2002 -0.3157 0.0723 -4.36 1.3e-05 ***
## Year2003 -0.3267 0.0721 -4.53 6.2e-06 ***
## Year2004 -0.3938 0.0710 -5.55 3.2e-08 ***
## Year2005 -0.4173 0.0714 -5.84 5.9e-09 ***
## Year2006 -0.3356 0.0709 -4.73 2.3e-06 ***
```

```

## Year2007          -0.3670      0.0695   -5.28  1.4e-07 ***
## Year2008          -0.3406      0.0691   -4.93  8.8e-07 ***
## Year2009          -0.3267      0.0697   -4.69  2.9e-06 ***
## Year2010          -0.3030      0.0684   -4.43  9.8e-06 ***
## Year2011          -0.3914      0.0689   -5.68  1.5e-08 ***
## Year2012          -0.3974      0.0700   -5.67  1.6e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.359
## Multiple R-squared:  0.07,   Adjusted R-squared:  0.0636
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 217 weights are ~= 1. The remaining 2268 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0291 0.8670 0.9500 0.8930 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.02e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.011
## Year            1.023 16          1.001

```

## Residuals from last author



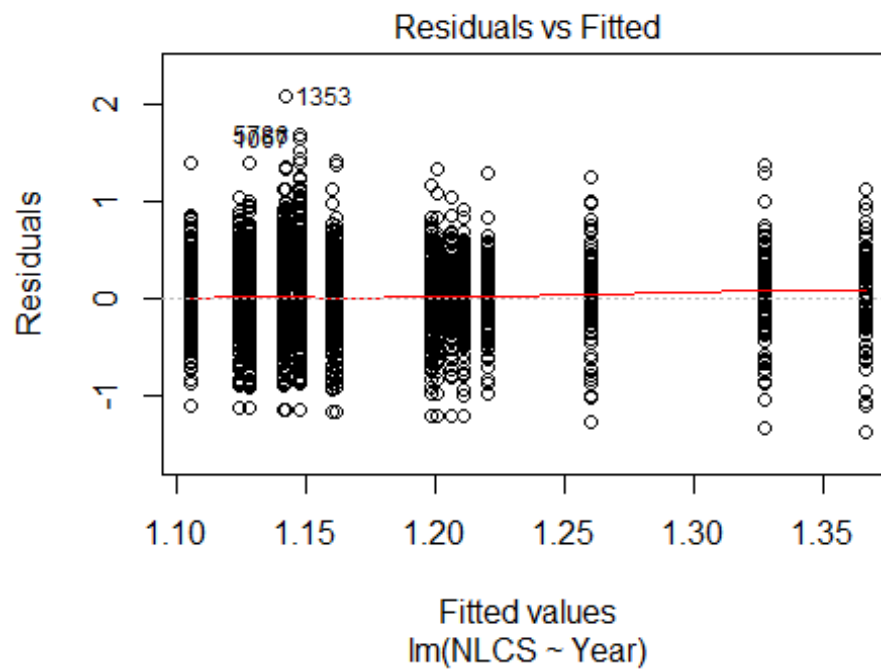
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4901 -0.2412 0.0042 0.2342 1.5241
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4901 0.0639 23.31 < 2e-16 ***
## LastAuthorFemale1 0.0208 0.0179 1.16 0.24494
## Year1997 -0.0410 0.0769 -0.53 0.59371
## Year1998 -0.2923 0.0699 -4.18 3.0e-05 ***
## Year1999 -0.3392 0.0712 -4.77 2.0e-06 ***
## Year2000 -0.2701 0.0719 -3.76 0.00018 ***
## Year2001 -0.3767 0.0696 -5.41 6.7e-08 ***
## Year2002 -0.3121 0.0725 -4.30 1.7e-05 ***
## Year2003 -0.3232 0.0722 -4.47 8.0e-06 ***
## Year2004 -0.3885 0.0711 -5.46 5.2e-08 ***
## Year2005 -0.4149 0.0715 -5.81 7.2e-09 ***
## Year2006 -0.3338 0.0707 -4.72 2.5e-06 ***
```

```

## Year2007          -0.3651      0.0695    -5.25  1.6e-07 ***
## Year2008          -0.3387      0.0692    -4.90  1.0e-06 ***
## Year2009          -0.3236      0.0697    -4.64  3.6e-06 ***
## Year2010          -0.2993      0.0684    -4.38  1.2e-05 ***
## Year2011          -0.3895      0.0691    -5.64  1.9e-08 ***
## Year2012          -0.3948      0.0700    -5.64  1.9e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.358
## Multiple R-squared:  0.0696, Adjusted R-squared:  0.0632
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 219 weights are ~= 1. The remaining 2266 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0309 0.8670 0.9510 0.8920 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.02e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2485"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2805"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 234 265 215 256 238 196 206 210 221 256 329 318 359 374 384
## 2011 2012
## 405 412
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 94 83 131 121 103 139 146 152 179 245 236 268 309 293
## 2011 2012

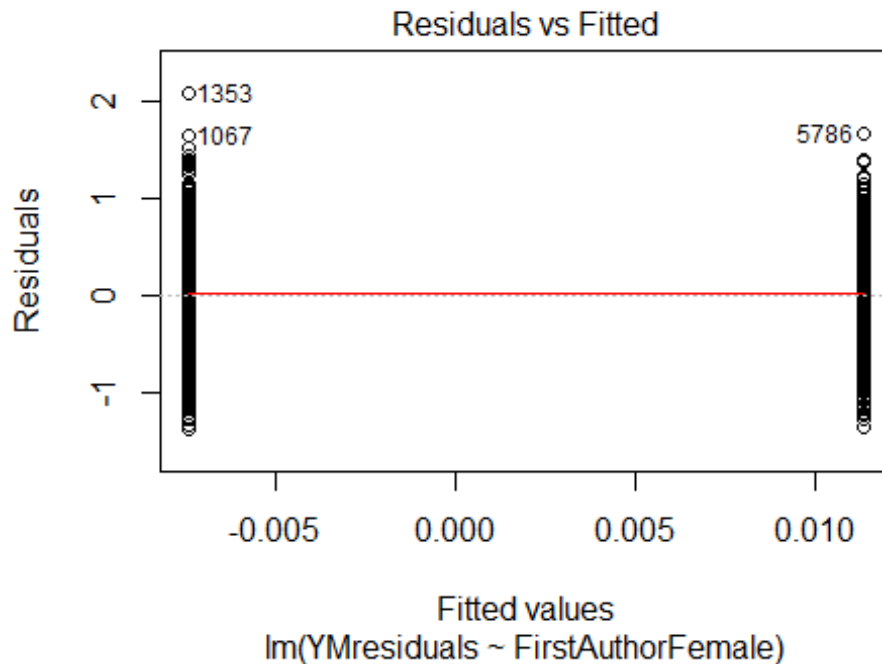
```

```
## 322 311
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 82 76 121 107 97 126 131 136 160 217 210 233 270 255
## 2011 2012
## 284 275
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 260, df = 16, p-value <2e-16
```



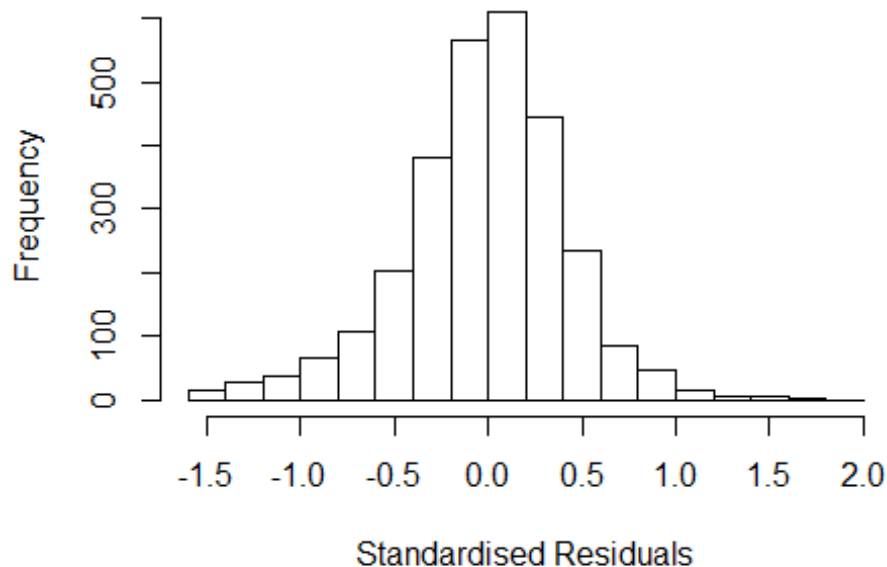
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 25, df = 1, p-value = 5e-07
```





```
## [1] "Female first author team size 2018 geometric mean: 3.39973222902795"
## [1] "Male first author team size 2018 geometric mean: 2.90450178460958"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 13000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.15938495445706"
## [1] "Male last author team size 2018 geometric mean: 3.10458741403864"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1      1.032
## LastAuthorFemale  1.065 1      1.032
## UniqueAuthors    1.337 4      1.037
## Year             1.379 16      1.010
```

## Residuals from first and last author and team size



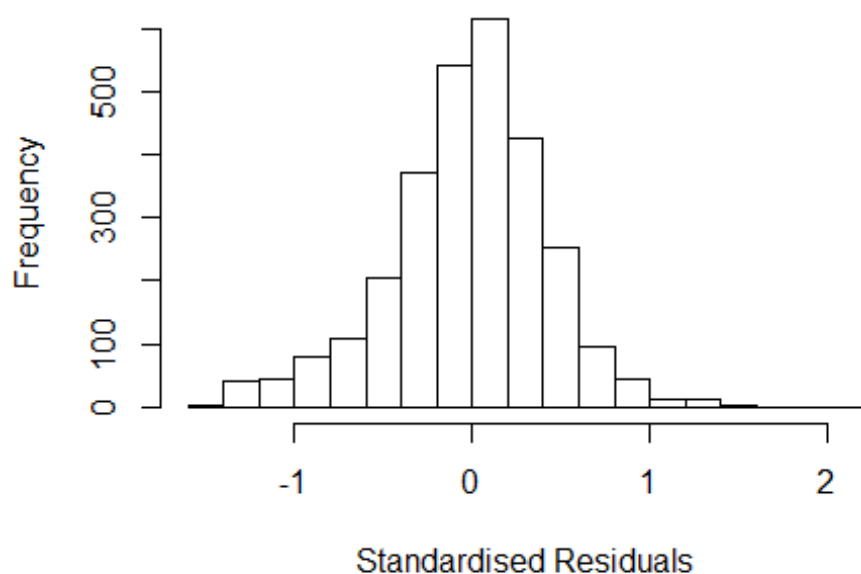
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.54270 -0.25077  0.00896  0.24422  1.97604
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0681    0.0526   20.30 < 2e-16 ***
## FirstAuthorFemale1 -0.0164    0.0154   -1.06  0.2882
## LastAuthorFemale1  0.0114    0.0170    0.67  0.5028
## UniqueAuthors2     0.1879    0.0306    6.13 9.9e-10 ***
## UniqueAuthors3     0.2387    0.0315    7.57 4.9e-14 ***
## UniqueAuthors4     0.2922    0.0336    8.70 < 2e-16 ***
## UniqueAuthors5     0.3477    0.0312   11.13 < 2e-16 ***
## Year1997           0.2270    0.0758    2.99  0.0028 **
## Year1998           0.1269    0.0817    1.55  0.1205
## Year1999           0.0884    0.1058    0.84  0.4034
```

```

## Year2000          -0.0160      0.1081   -0.15   0.8823
## Year2001           0.0394      0.0670    0.59   0.5569
## Year2002          -0.0667      0.0587   -1.14   0.2556
## Year2003          -0.0891      0.0596   -1.49   0.1351
## Year2004          -0.0852      0.0598   -1.42   0.1543
## Year2005          -0.0815      0.0571   -1.43   0.1536
## Year2006          -0.1862      0.0555   -3.36   0.0008 ***
## Year2007          -0.1371      0.0542   -2.53   0.0114 *
## Year2008          -0.1387      0.0540   -2.57   0.0103 *
## Year2009          -0.1534      0.0538   -2.85   0.0044 **
## Year2010          -0.1642      0.0547   -3.00   0.0027 **
## Year2011          -0.1394      0.0540   -2.58   0.0098 **
## Year2012          -0.1611      0.0553   -2.91   0.0036 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.375
## Multiple R-squared:  0.101, Adjusted R-squared:  0.094
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## observation 432 is an outlier with |weight| = 0 ( < 3.5e-05);
## 226 weights are ~= 1. The remaining 2634 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0157 0.8610 0.9520 0.8830 0.9870 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.50e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.059 1 1.029
## LastAuthorFemale 1.046 1 1.023
## Year 1.078 16 1.002

```

## Residuals from first and last author



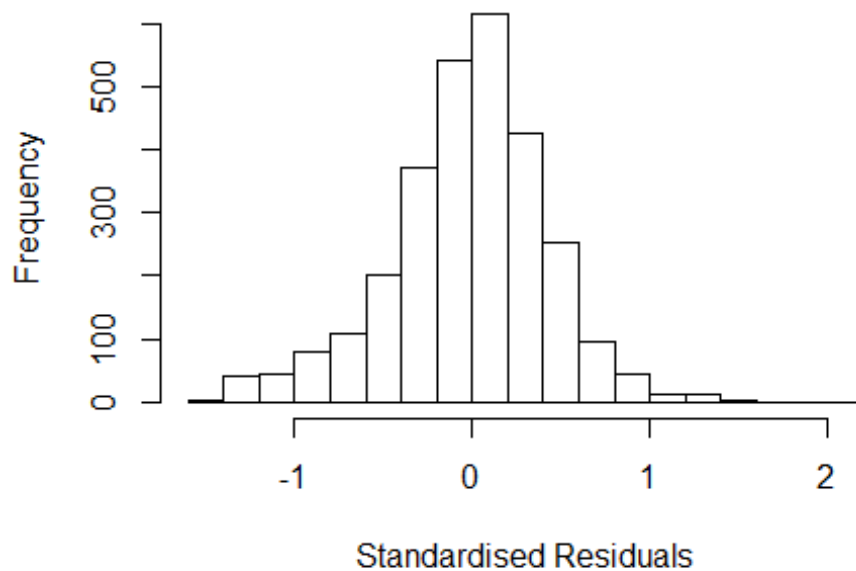
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4480 -0.2600 0.0114 0.2514 2.0208
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22410 0.04889 25.04 <2e-16 ***
## FirstAuthorFemale1 0.00988 0.01596 0.62 0.5361
## LastAuthorFemale1 -0.00116 0.01719 -0.07 0.9461
## Year1997 0.22390 0.07458 3.00 0.0027 **
## Year1998 0.12168 0.08215 1.48 0.1386
## Year1999 0.03744 0.11581 0.32 0.7465
## Year2000 -0.02888 0.10806 -0.27 0.7893
## Year2001 0.10670 0.06750 1.58 0.1141
## Year2002 0.01051 0.05851 0.18 0.8574
## Year2003 -0.01279 0.05868 -0.22 0.8274
## Year2004 -0.02238 0.05949 -0.38 0.7068
## Year2005 -0.00391 0.05699 -0.07 0.9453
```

```

## Year2006          -0.12322      0.05522      -2.23      0.0257 *
## Year2007          -0.06219      0.05360      -1.16      0.2460
## Year2008          -0.06619      0.05331      -1.24      0.2145
## Year2009          -0.08447      0.05301      -1.59      0.1112
## Year2010          -0.08325      0.05416      -1.54      0.1244
## Year2011          -0.06714      0.05346      -1.26      0.2093
## Year2012          -0.08588      0.05525      -1.55      0.1202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0293, Adjusted R-squared:  0.0232
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## observation 432 is an outlier with |weight| = 0 ( < 3.5e-05);
## 227 weights are ~= 1. The remaining 2633 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0194 0.8610 0.9520 0.8840 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.50e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1          1.022
## Year              1.044 16          1.001

```

## Residuals from first author



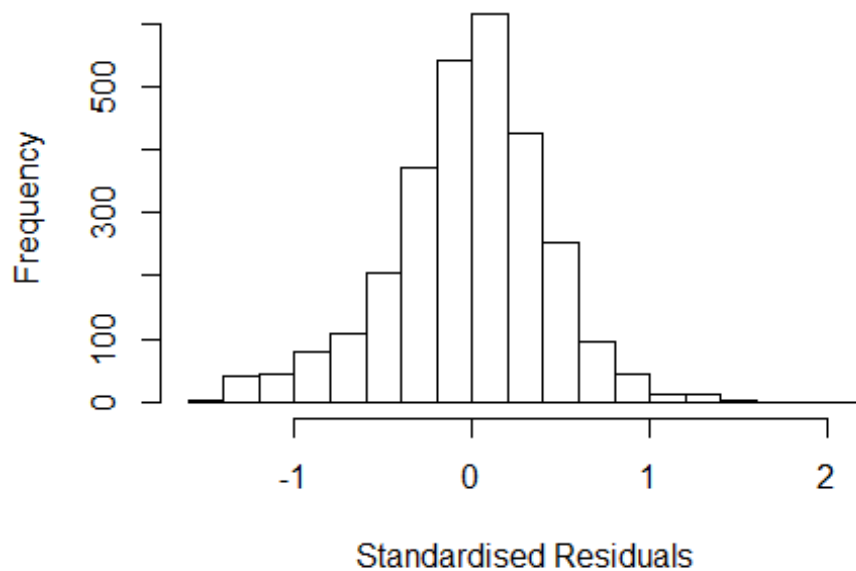
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4477 -0.2597 0.0116 0.2517 2.0214
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22372 0.04864 25.16 <2e-16 ***
## FirstAuthorFemale1 0.00969 0.01589 0.61 0.5421
## Year1997 0.22397 0.07441 3.01 0.0026 **
## Year1998 0.12170 0.08219 1.48 0.1388
## Year1999 0.03702 0.11569 0.32 0.7490
## Year2000 -0.02915 0.10796 -0.27 0.7872
## Year2001 0.10683 0.06744 1.58 0.1133
## Year2002 0.01062 0.05852 0.18 0.8560
## Year2003 -0.01271 0.05869 -0.22 0.8286
## Year2004 -0.02221 0.05943 -0.37 0.7086
## Year2005 -0.00383 0.05700 -0.07 0.9465
## Year2006 -0.12314 0.05522 -2.23 0.0258 *
```

```

## Year2007          -0.06211      0.05360      -1.16      0.2466
## Year2008          -0.06609      0.05329      -1.24      0.2150
## Year2009          -0.08431      0.05301      -1.59      0.1119
## Year2010          -0.08323      0.05416      -1.54      0.1245
## Year2011          -0.06703      0.05343      -1.25      0.2098
## Year2012          -0.08579      0.05525      -1.55      0.1206
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.0292, Adjusted R-squared:  0.0234
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## observation 432 is an outlier with |weight| = 0 ( < 3.5e-05);
## 229 weights are ~= 1. The remaining 2631 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0199 0.8620 0.9520 0.8840 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.033 1          1.017
## Year          1.033 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4518 -0.2615 0.0117 0.2502 2.0188
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.227273 0.048577 25.26 <2e-16 ***
## LastAuthorFemale1 0.000877 0.017124 0.05 0.9592
## Year1997 0.224535 0.074543 3.01 0.0026 **
## Year1998 0.121309 0.082277 1.47 0.1405
## Year1999 0.037726 0.115601 0.33 0.7442
## Year2000 -0.030053 0.108142 -0.28 0.7811
## Year2001 0.106386 0.067478 1.58 0.1150
## Year2002 0.011222 0.058518 0.19 0.8479
## Year2003 -0.013219 0.058713 -0.23 0.8219
## Year2004 -0.021946 0.059517 -0.37 0.7124
## Year2005 -0.003067 0.056927 -0.05 0.9570
## Year2006 -0.123140 0.055244 -2.23 0.0259 *
```

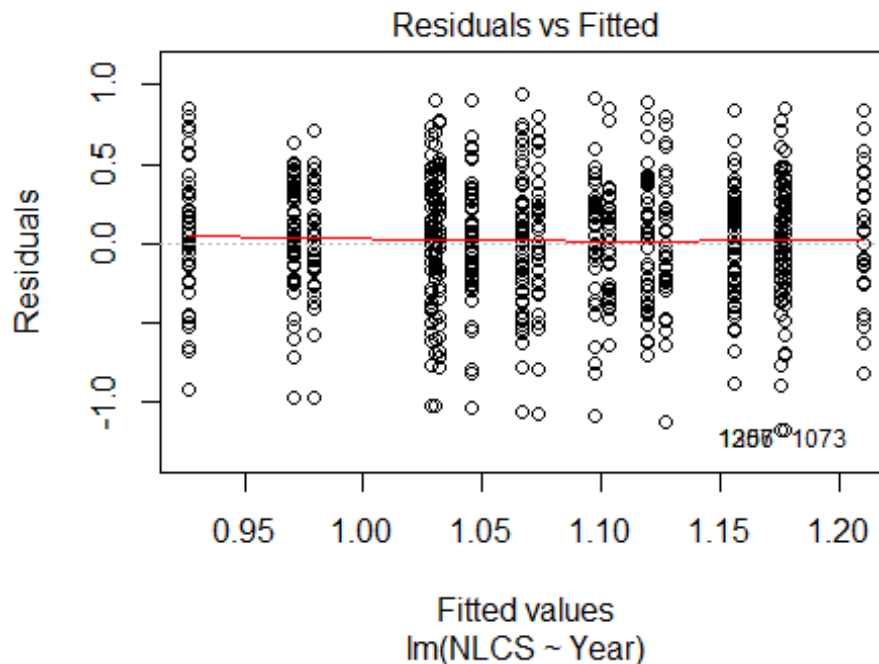


```

## Year2007      -0.062228    0.053612   -1.16    0.2459
## Year2008      -0.065743    0.053302   -1.23    0.2175
## Year2009      -0.084323    0.053044   -1.59    0.1120
## Year2010      -0.082222    0.054140   -1.52    0.1290
## Year2011      -0.066428    0.053464   -1.24    0.2142
## Year2012      -0.084731    0.055188   -1.54    0.1248
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.385
## Multiple R-squared:  0.0292, Adjusted R-squared:  0.0234
## Convergence in 24 IRWLS iterations
##
## Robustness weights:
## observation 432 is an outlier with |weight| = 0 ( < 3.5e-05);
## 231 weights are ~= 1. The remaining 2629 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0176 0.8610 0.9510 0.8830 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2861"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2806"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   66   77   65   56   72   56   56   38   58   84   76   64   63   75
## 2011 2012
##   67   69
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   36   46   34   34   46   35   34   31   40   66   56   49   51   60

```

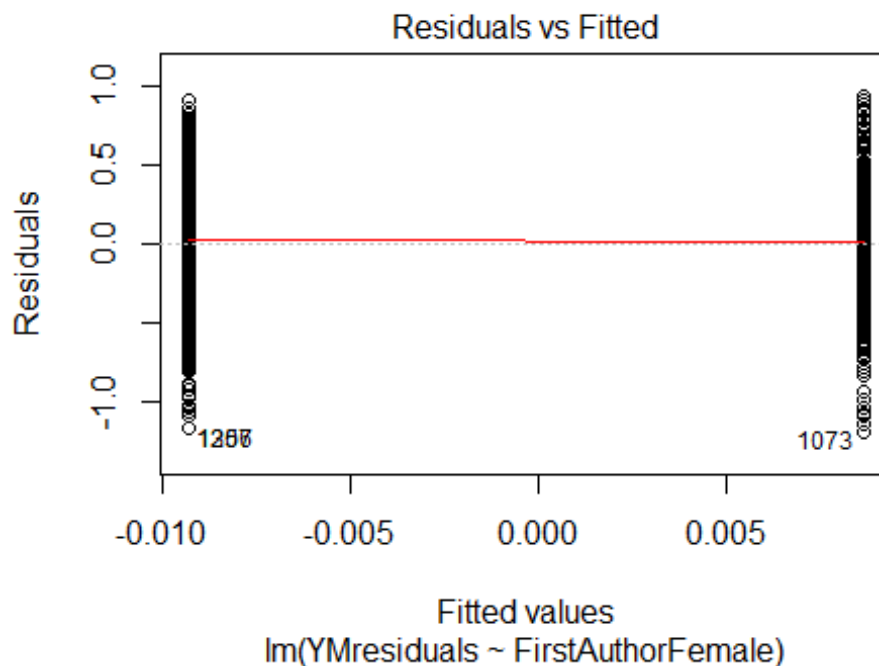
```
## 2011 2012
## 53 58
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 26 33 41 32 32 43 33 28 28 36 60 48 43 46 54
## 2011 2012
## 44 51
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 12, df = 16, p-value = 0.7
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07
## [1] "Female first author team size 2018 geometric mean: 4.29109355007577"
## [1] "Male first author team size 2018 geometric mean: 3.75056800945124"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.43603171111633"
## [1] "Male last author team size 2018 geometric mean: 3.53411404822032"

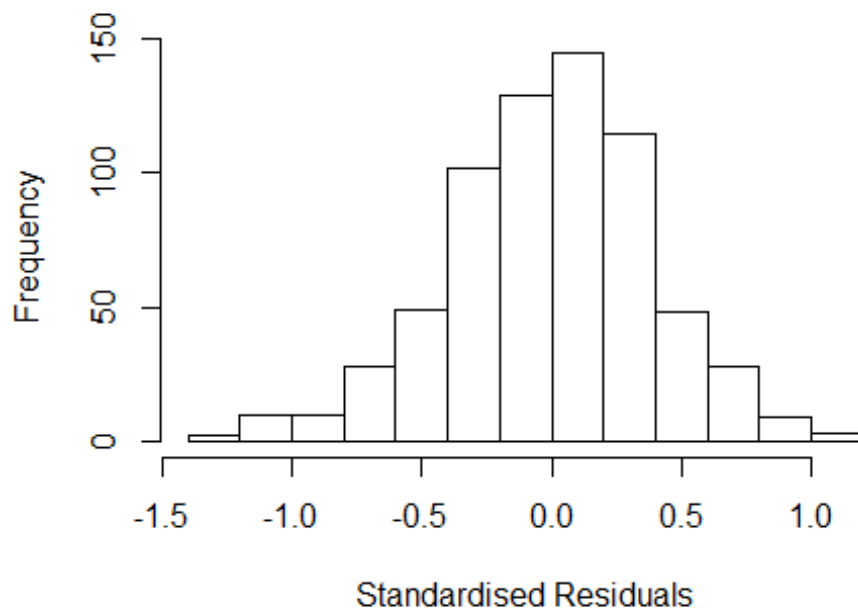
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.191	1	1.091
LastAuthorFemale	1.230	1	1.109
UniqueAuthors	1.764	4	1.074
Year	1.981	16	1.022

## Residuals from first and last author and team size



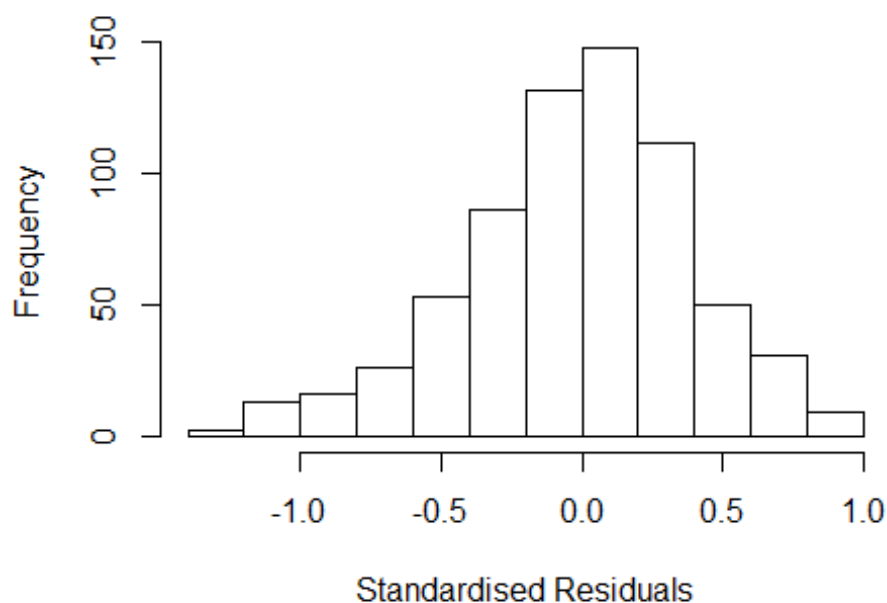
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2535 -0.2624 0.0138 0.2453 1.1512
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0013 0.0974 10.28 < 2e-16 ***
## FirstAuthorFemale1 -0.0137 0.0322 -0.43 0.66916
## LastAuthorFemale1 -0.0324 0.0339 -0.96 0.33989
## UniqueAuthors2 0.2381 0.0658 3.62 0.00032 ***
## UniqueAuthors3 0.2655 0.0677 3.92 9.8e-05 ***
## UniqueAuthors4 0.2884 0.0679 4.25 2.5e-05 ***
## UniqueAuthors5 0.3505 0.0635 5.52 4.9e-08 ***
## Year1997 -0.1059 0.1044 -1.01 0.31072
## Year1998 -0.0994 0.1027 -0.97 0.33361
## Year1999 -0.1980 0.1042 -1.90 0.05788 .
```

```

## Year2000          -0.1295      0.0992    -1.30   0.19237
## Year2001          -0.2890      0.1070    -2.70   0.00711 **
## Year2002          -0.2105      0.0950    -2.21   0.02711 *
## Year2003          -0.1327      0.1126    -1.18   0.23935
## Year2004          -0.1482      0.1146    -1.29   0.19664
## Year2005          -0.2179      0.1064    -2.05   0.04101 *
## Year2006          -0.1551      0.0900    -1.72   0.08515 .
## Year2007          -0.2016      0.0981    -2.06   0.04016 *
## Year2008          -0.0428      0.0935    -0.46   0.64757
## Year2009          -0.1994      0.1005    -1.98   0.04769 *
## Year2010          -0.1114      0.0896    -1.24   0.21414
## Year2011          -0.0134      0.0967    -0.14   0.89016
## Year2012          -0.1604      0.0992    -1.62   0.10638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.376
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0742
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 617 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.242  0.873  0.950  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.47e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.121 1 1.059
## LastAuthorFemale 1.157 1 1.076
## Year 1.177 16 1.005

```

## Residuals from first and last author



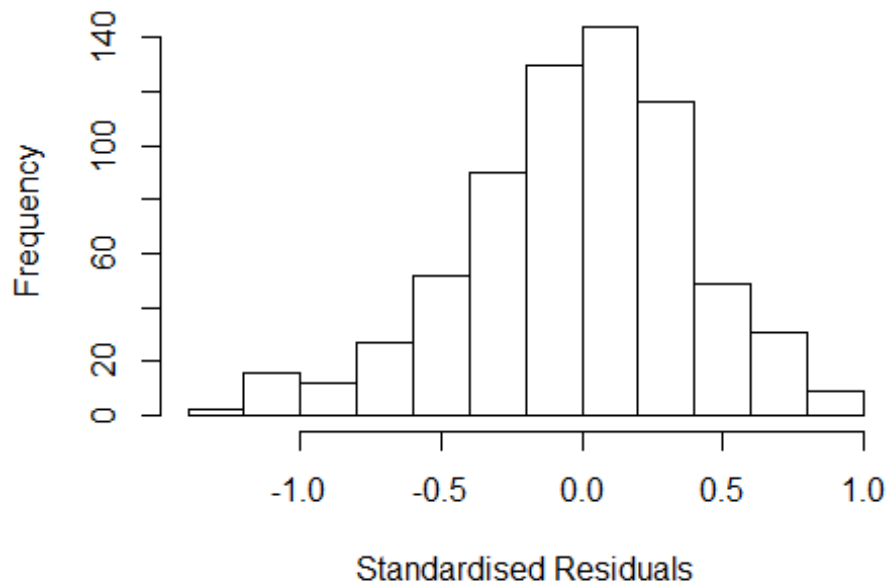
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2673 -0.2575 0.0106 0.2560 0.9227
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.24291 0.08712 14.27 <2e-16 ***
## FirstAuthorFemale1 0.00942 0.03246 0.29 0.772
## LastAuthorFemale1 -0.04252 0.03385 -1.26 0.210
## Year1997 -0.13150 0.11409 -1.15 0.250
## Year1998 -0.12554 0.10784 -1.16 0.245
## Year1999 -0.21791 0.10813 -2.02 0.044 *
## Year2000 -0.13099 0.10398 -1.26 0.208
## Year2001 -0.29138 0.11843 -2.46 0.014 *
## Year2002 -0.19728 0.10335 -1.91 0.057 .
## Year2003 -0.11631 0.11633 -1.00 0.318
## Year2004 -0.11023 0.11725 -0.94 0.347
## Year2005 -0.20046 0.11710 -1.71 0.087 .
```

```

## Year2006      -0.16589    0.09629   -1.72    0.085 .
## Year2007      -0.22607    0.10378   -2.18    0.030 *
## Year2008      -0.03975    0.10323   -0.39    0.700
## Year2009      -0.20237    0.10664   -1.90    0.058 .
## Year2010      -0.08795    0.09700   -0.91    0.365
## Year2011       0.02438    0.09970    0.24    0.807
## Year2012      -0.15298    0.10700   -1.43    0.153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.383
## Multiple R-squared:  0.0418, Adjusted R-squared:  0.0156
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 58 weights are ~= 1. The remaining 620 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.252  0.870   0.953   0.898   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.47e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.074 1      1.036
## Year              1.074 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2563 -0.2554 0.0137 0.2475 0.9251
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.237526 0.087767 14.10 <2e-16 ***
## FirstAuthorFemale1 0.000771 0.031770 0.02 0.981
## Year1997 -0.131907 0.114687 -1.15 0.250
## Year1998 -0.125629 0.109252 -1.15 0.251
## Year1999 -0.213240 0.108262 -1.97 0.049 *
## Year2000 -0.136531 0.104135 -1.31 0.190
## Year2001 -0.295068 0.118692 -2.49 0.013 *
## Year2002 -0.200476 0.104171 -1.92 0.055 .
## Year2003 -0.122536 0.115860 -1.06 0.291
## Year2004 -0.116049 0.118704 -0.98 0.329
## Year2005 -0.204833 0.116288 -1.76 0.079 .
## Year2006 -0.175286 0.096201 -1.82 0.069 .
```

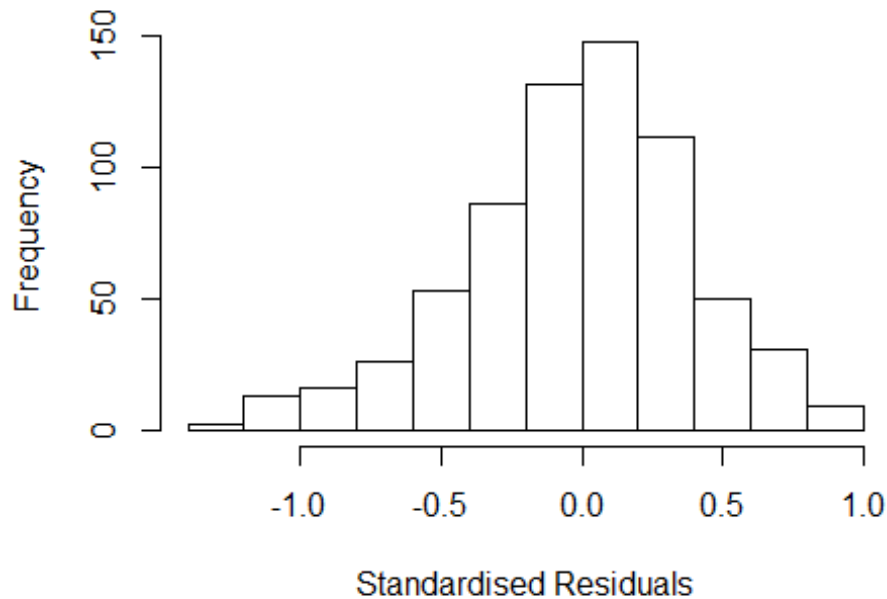


```

## Year2007          -0.235687    0.103510    -2.28    0.023 *
## Year2008          -0.044791    0.103750    -0.43    0.666
## Year2009          -0.210676    0.106718    -1.97    0.049 *
## Year2010          -0.093926    0.097629    -0.96    0.336
## Year2011           0.018758    0.100192     0.19    0.852
## Year2012          -0.157433    0.107042    -1.47    0.142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.0395, Adjusted R-squared:  0.0148
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 624 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.262  0.867   0.953   0.899   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.47e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.106 1          1.051
## Year              1.106 16          1.003

```

## Residuals from last author



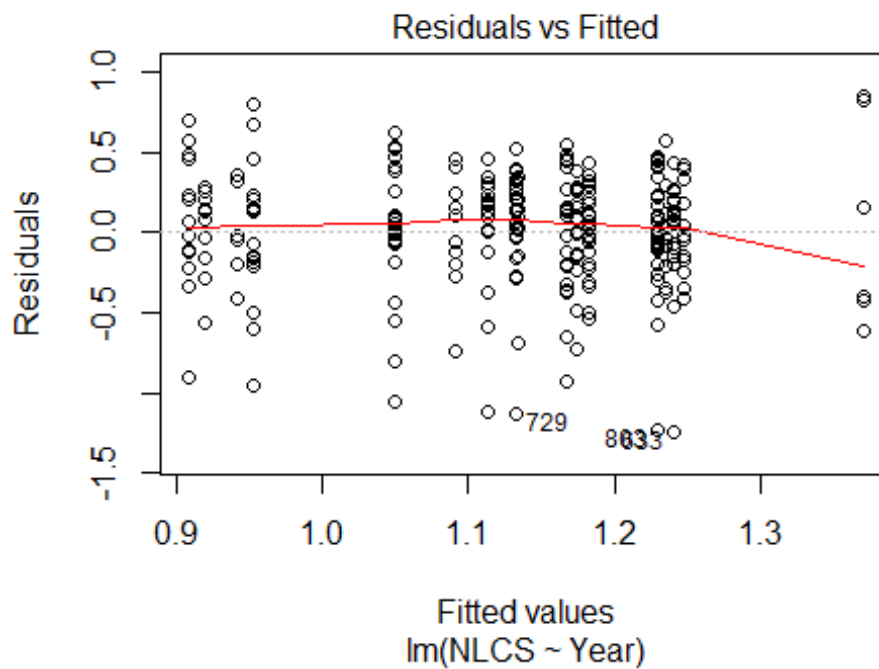
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2721 -0.2590 0.0147 0.2588 0.9253
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2468 0.0849 14.68 <2e-16 ***
## LastAuthorFemale1 -0.0404 0.0331 -1.22 0.223
## Year1997 -0.1317 0.1138 -1.16 0.247
## Year1998 -0.1247 0.1081 -1.15 0.249
## Year1999 -0.2178 0.1078 -2.02 0.044 *
## Year2000 -0.1306 0.1041 -1.25 0.210
## Year2001 -0.2914 0.1183 -2.46 0.014 *
## Year2002 -0.1955 0.1033 -1.89 0.059 .
## Year2003 -0.1153 0.1163 -0.99 0.322
## Year2004 -0.1088 0.1177 -0.92 0.356
## Year2005 -0.2000 0.1172 -1.71 0.088 .
## Year2006 -0.1657 0.0962 -1.72 0.085 .
```

```

## Year2007          -0.2253      0.1036    -2.17      0.030 *
## Year2008          -0.0406      0.1027    -0.39      0.693
## Year2009          -0.2025      0.1066    -1.90      0.058 .
## Year2010          -0.0865      0.0972    -0.89      0.374
## Year2011           0.0253      0.0996      0.25      0.800
## Year2012          -0.1530      0.1069    -1.43      0.153
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared:  0.0416, Adjusted R-squared:  0.0169
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 60 weights are ~= 1. The remaining 618 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.246  0.868  0.953  0.897  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.47e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 678"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2807"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32   34   27   27   38   39   30   24   35   33   33   38   41   52   40
## 2011 2012
##   55   51
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   12    8    6    9   12   15   10   16   15   25   19   25   19   18
## 2011 2012

```

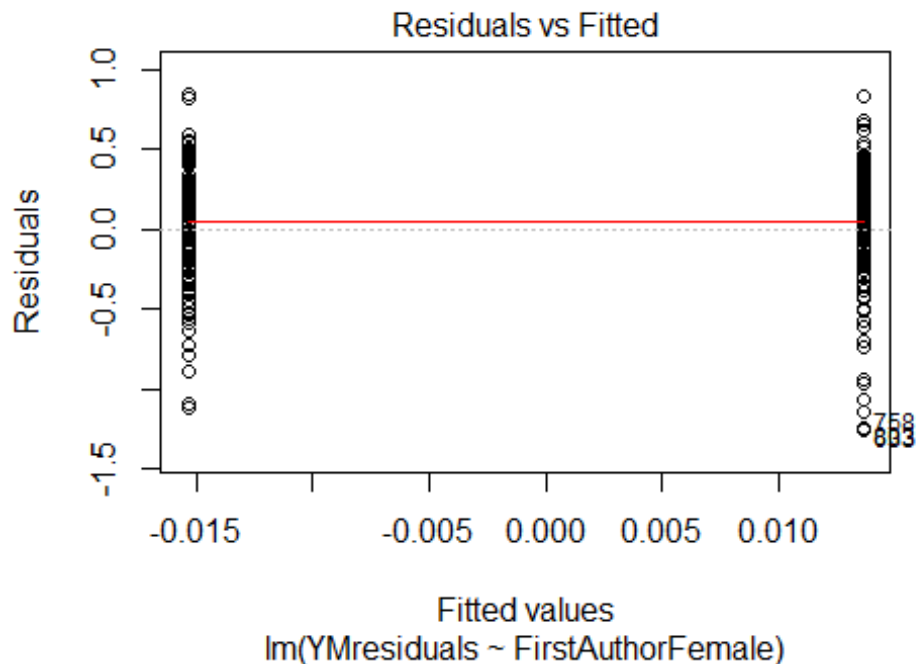
```
## 26 26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 7 10 8 3 5 8 9 7 15 13 21 16 19 15 12
## 2011 2012
## 25 22
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 21, df = 16, p-value = 0.2
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.029, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 3.61489122200056"
## [1] "Male first author team size 2018 geometric mean: 4.82615548036294"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 41, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.7484362649004"
## [1] "Male last author team size 2018 geometric mean: 4.01070682111648"

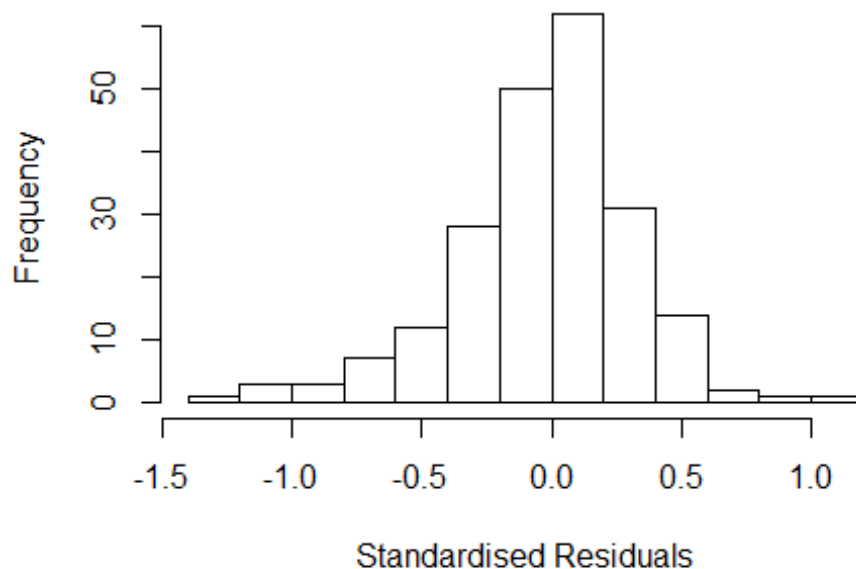
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 64, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.486	1	1.219
LastAuthorFemale	1.958	1	1.399
UniqueAuthors	5.755	4	1.245
Year	8.119	16	1.068

## Residuals from first and last author and team size



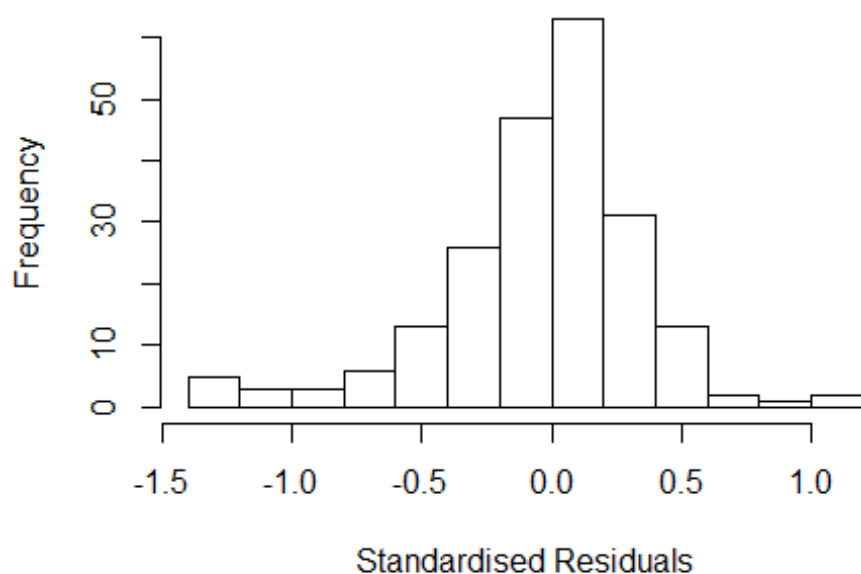
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23062 -0.20440  0.00542  0.19114  1.02464
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9000    0.2661    3.38 0.00087 ***
## FirstAuthorFemale1  0.0775    0.0508    1.53 0.12856
## LastAuthorFemale1 -0.0291    0.0630   -0.46 0.64430
## UniqueAuthors2    0.0497    0.1574    0.32 0.75255
## UniqueAuthors3    0.2026    0.1471    1.38 0.17008
## UniqueAuthors4    0.3970    0.1620    2.45 0.01514 *
## UniqueAuthors5    0.2744    0.1461    1.88 0.06193 .
## Year1997          0.1720    0.2555    0.67 0.50157
## Year1998         -0.0487    0.2581   -0.19 0.85059
## Year1999         -0.4426    0.2630   -1.68 0.09402 .
```

```

## Year2000          0.1718      0.2692      0.64  0.52419
## Year2001         -0.1661      0.2572     -0.65  0.51915
## Year2002         -0.2213      0.2976     -0.74  0.45790
## Year2003         -0.1502      0.2550     -0.59  0.55660
## Year2004         -0.1410      0.2427     -0.58  0.56178
## Year2005          0.0724      0.2515      0.29  0.77368
## Year2006          0.0276      0.2388      0.12  0.90815
## Year2007          0.0107      0.2536      0.04  0.96624
## Year2008         -0.1534      0.2402     -0.64  0.52383
## Year2009          0.1018      0.2317      0.44  0.66085
## Year2010         -0.0097      0.2533     -0.04  0.96948
## Year2011          0.0796      0.2402      0.33  0.74062
## Year2012          0.1335      0.2399      0.56  0.57865
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.32
## Multiple R-squared:  0.248, Adjusted R-squared:  0.162
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 196 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.868  0.957  0.892  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.65e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.342 1      1.159
## LastAuthorFemale  1.766 1      1.329
## Year              2.252 16      1.026

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3478 -0.2105  0.0245  0.1806  1.0860
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11305    0.29353   3.79   0.0002 ***
## FirstAuthorFemale1 0.06553    0.04850   1.35   0.1782
## LastAuthorFemale1 0.00369    0.05840   0.06   0.9497
## Year1997         0.13672    0.30960   0.44   0.6593
## Year1998         0.01719    0.31550   0.05   0.9566
## Year1999        -0.43355    0.31340  -1.38   0.1681
## Year2000         0.11880    0.31866   0.37   0.7097
## Year2001        -0.16208    0.31441  -0.52   0.6068
## Year2002        -0.17218    0.40961  -0.42   0.6747
## Year2003        -0.09959    0.32253  -0.31   0.7578
## Year2004        -0.22790    0.32034  -0.71   0.4777
## Year2005         0.04793    0.31773   0.15   0.8802
```

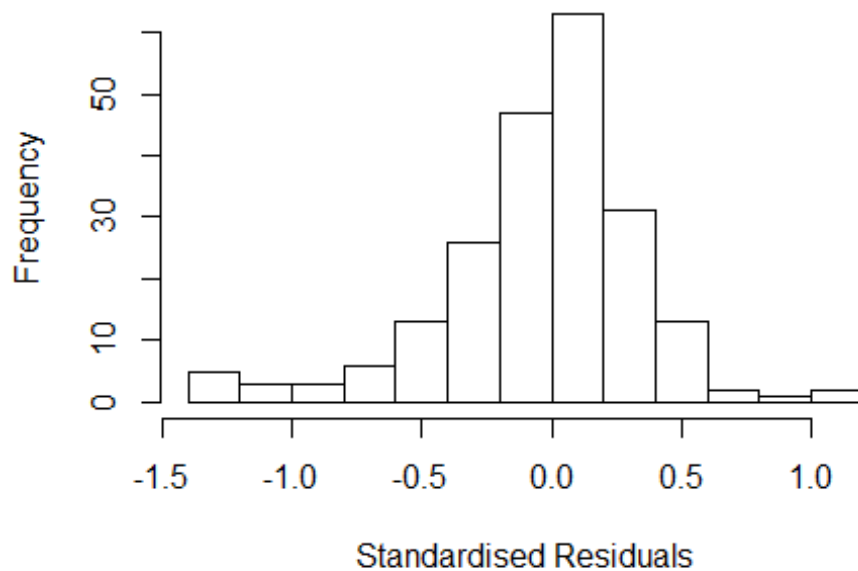


```

## Year2006          0.03021    0.30605    0.10    0.9215
## Year2007          0.07189    0.32696    0.22    0.8262
## Year2008         -0.12984    0.31162   -0.42    0.6774
## Year2009          0.15081    0.30781    0.49    0.6247
## Year2010          0.05911    0.31839    0.19    0.8529
## Year2011          0.11678    0.31177    0.37    0.7084
## Year2012          0.16549    0.30273    0.55    0.5852
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.312
## Multiple R-squared:  0.163, Adjusted R-squared:  0.0862
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0225 0.8620 0.9570 0.8720 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.65e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.311 1      1.145
## Year              1.311 16      1.008

```

## Residuals from first author



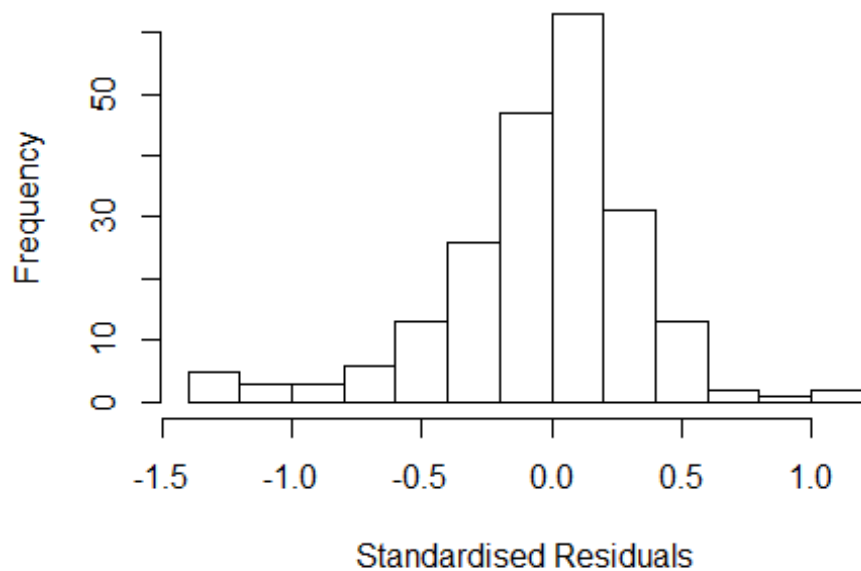
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3452 -0.2091 0.0237 0.1807 1.0858
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1132 0.2942 3.78 0.0002 ***
## FirstAuthorFemale1 0.0657 0.0481 1.37 0.1730
## Year1997 0.1388 0.3080 0.45 0.6527
## Year1998 0.0186 0.3150 0.06 0.9529
## Year1999 -0.4326 0.3141 -1.38 0.1699
## Year2000 0.1204 0.3176 0.38 0.7050
## Year2001 -0.1623 0.3148 -0.52 0.6068
## Year2002 -0.1712 0.4340 -0.39 0.6936
## Year2003 -0.0998 0.3230 -0.31 0.7577
## Year2004 -0.2275 0.3213 -0.71 0.4798
## Year2005 0.0483 0.3183 0.15 0.8795
## Year2006 0.0309 0.3065 0.10 0.9197
```

```

## Year2007          0.0730      0.3276      0.22      0.8239
## Year2008         -0.1297      0.3124     -0.42      0.6786
## Year2009          0.1524      0.3081      0.49      0.6213
## Year2010          0.0609      0.3177      0.19      0.8483
## Year2011          0.1182      0.3118      0.38      0.7050
## Year2012          0.1663      0.3036      0.55      0.5845
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.312
## Multiple R-squared:  0.163, Adjusted R-squared:  0.0908
## Convergence in 26 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0236 0.8630 0.9580 0.8720 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.65e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.677 1          1.295
## Year              1.677 16          1.016

```

## Residuals from last author



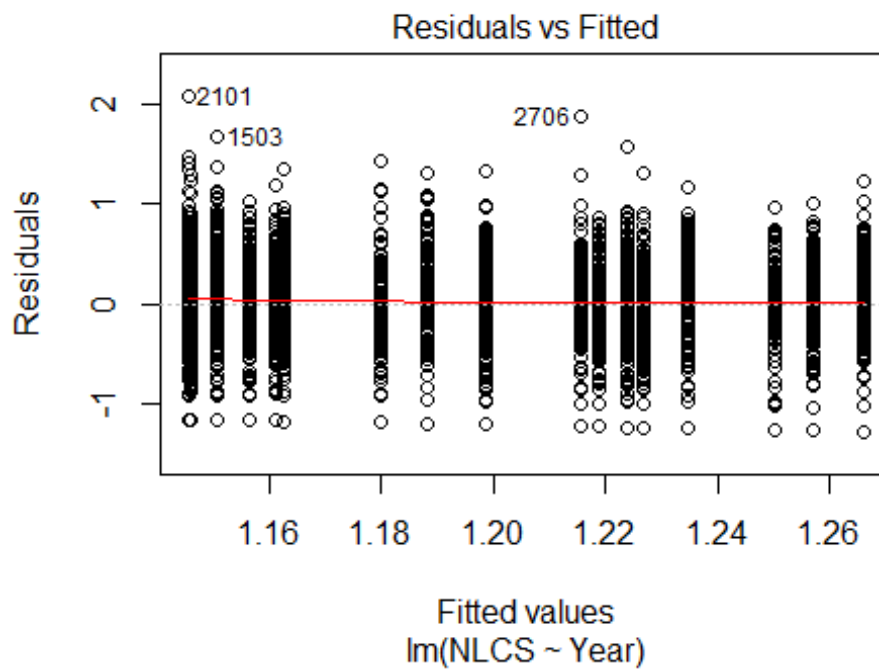
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31204 -0.19629 0.00421 0.18235 1.09714
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12386 0.28457 3.95 0.00011 ***
## LastAuthorFemale1 0.00697 0.05781 0.12 0.90410
## Year1997 0.16305 0.29648 0.55 0.58299
## Year1998 0.04027 0.30462 0.13 0.89497
## Year1999 -0.40312 0.29898 -1.35 0.17910
## Year2000 0.11593 0.30476 0.38 0.70406
## Year2001 -0.14273 0.30063 -0.47 0.63549
## Year2002 -0.16988 0.42252 -0.40 0.68807
## Year2003 -0.09294 0.31646 -0.29 0.76930
## Year2004 -0.19737 0.30834 -0.64 0.52285
## Year2005 0.08315 0.30392 0.27 0.78470
## Year2006 0.06095 0.29289 0.21 0.83538
```

```

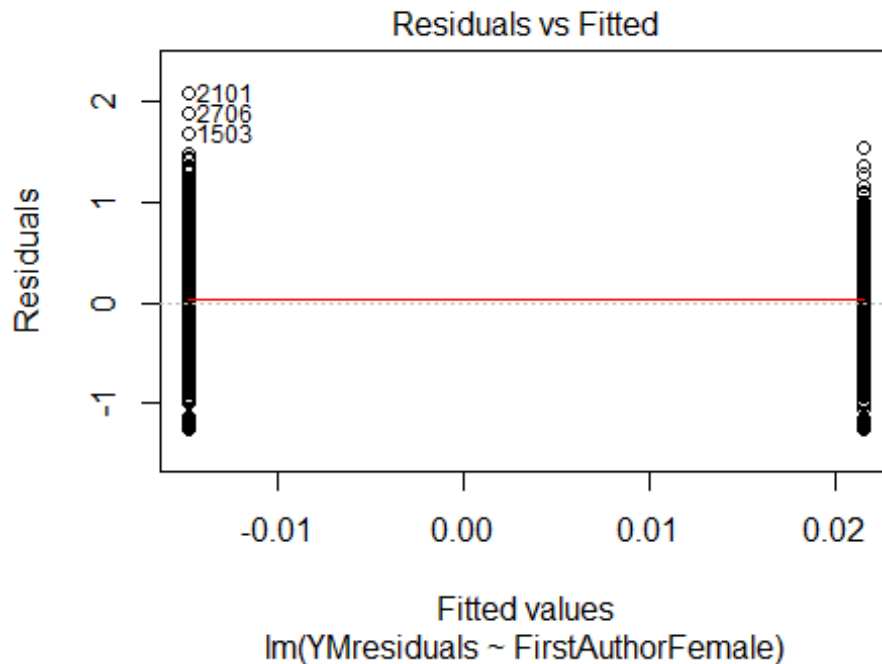
## Year2007      0.08219      0.31737      0.26  0.79594
## Year2008     -0.11139      0.30126     -0.37  0.71196
## Year2009      0.17095      0.29534      0.58  0.56336
## Year2010      0.08802      0.30415      0.29  0.77258
## Year2011      0.14196      0.29827      0.48  0.63464
## Year2012      0.18121      0.29146      0.62  0.53484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.315
## Multiple R-squared:  0.154, Adjusted R-squared:  0.0812
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~= 1. The remaining 191 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0431 0.8620 0.9580 0.8690 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.65e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 215"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2808"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 312 373 361 425 413 341 351 327 340 345 383 433 423 418 436
## 2011 2012
## 479 469
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 110 131 124 211 181 141 180 186 217 223 234 273 262 289 286
## 2011 2012

```

```
## 333 314
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 97 110 109 181 149 132 154 150 172 172 185 235 222 253 261
## 2011 2012
## 299 278
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 200, df = 16, p-value <2e-16
```

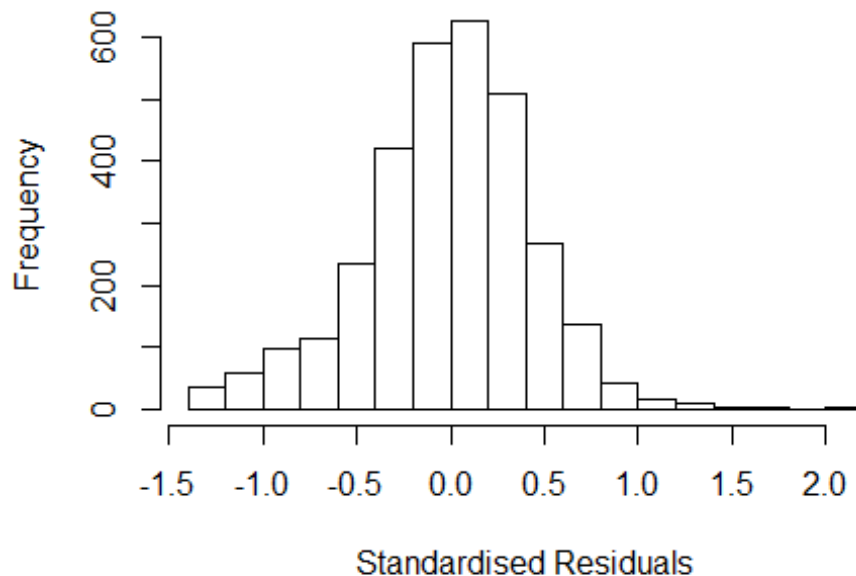


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 16, df = 1, p-value = 5e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 4.60939415279541"
## [1] "Male first author team size 2018 geometric mean: 4.44313099531186"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9200, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.28002109272138"
## [1] "Male last author team size 2018 geometric mean: 4.66048455937345"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7600, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1          1.019
## LastAuthorFemale  1.044 1          1.022
## UniqueAuthors    1.146 4          1.017
## Year              1.198 16         1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3608 -0.2708 0.0104 0.2696 2.1705
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0375 0.0519 20.00 < 2e-16 ***
## FirstAuthorFemale1 0.0118 0.0155 0.76 0.44852
## LastAuthorFemale1 0.0129 0.0176 0.74 0.46167
## UniqueAuthors2 0.2155 0.0391 5.52 3.7e-08 ***
## UniqueAuthors3 0.2636 0.0380 6.93 5.0e-12 ***
## UniqueAuthors4 0.3233 0.0375 8.61 < 2e-16 ***
## UniqueAuthors5 0.4007 0.0349 11.47 < 2e-16 ***
## Year1997 -0.0354 0.0664 -0.53 0.59406
## Year1998 -0.1250 0.0687 -1.82 0.06896 .
## Year1999 -0.1010 0.0702 -1.44 0.15053
```

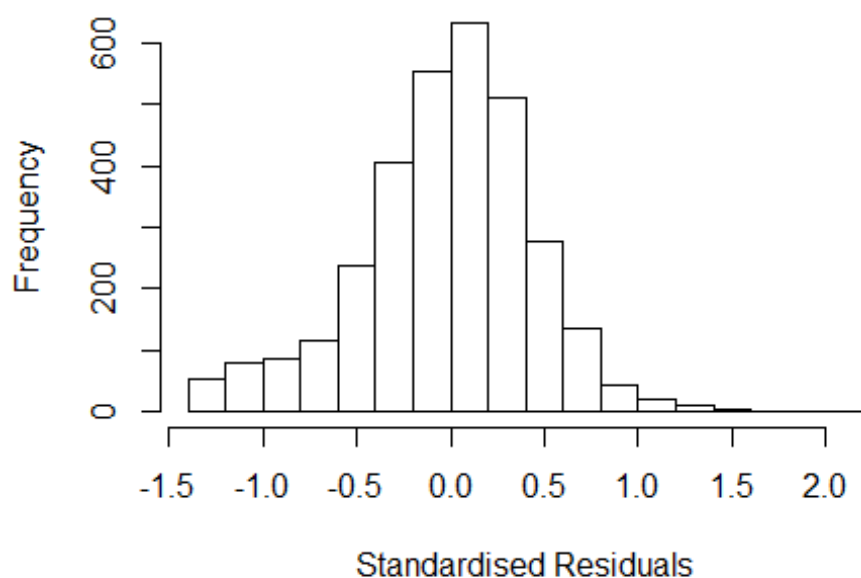


```

## Year2000          -0.1530      0.0674    -2.27  0.02332 *
## Year2001          -0.1083      0.0627    -1.73  0.08390 .
## Year2002          -0.1160      0.0557    -2.08  0.03717 *
## Year2003          -0.0671      0.0534    -1.26  0.20893
## Year2004          -0.1152      0.0513    -2.25  0.02480 *
## Year2005          -0.1203      0.0513    -2.35  0.01900 *
## Year2006          -0.0987      0.0499    -1.98  0.04789 *
## Year2007          -0.1595      0.0503    -3.17  0.00152 **
## Year2008          -0.1565      0.0505    -3.10  0.00195 **
## Year2009          -0.1675      0.0510    -3.29  0.00103 **
## Year2010          -0.1206      0.0501    -2.41  0.01606 *
## Year2011          -0.1655      0.0498    -3.32  0.00091 ***
## Year2012          -0.1804      0.0524    -3.44  0.00059 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.397
## Multiple R-squared:  0.0848, Adjusted R-squared:  0.0784
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 2 observations c(597,793) are outliers with |weight| = 0 ( < 3.2e-05);
## 272 weights are ~= 1. The remaining 2885 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0395 0.8640 0.9480 0.8870 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.17e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500          50          2          1          1000          200
## trace.lev mts compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## LastAuthorFemale 1.045 1          1.022
## Year 1.084 16          1.003

```

## Residuals from first and last author



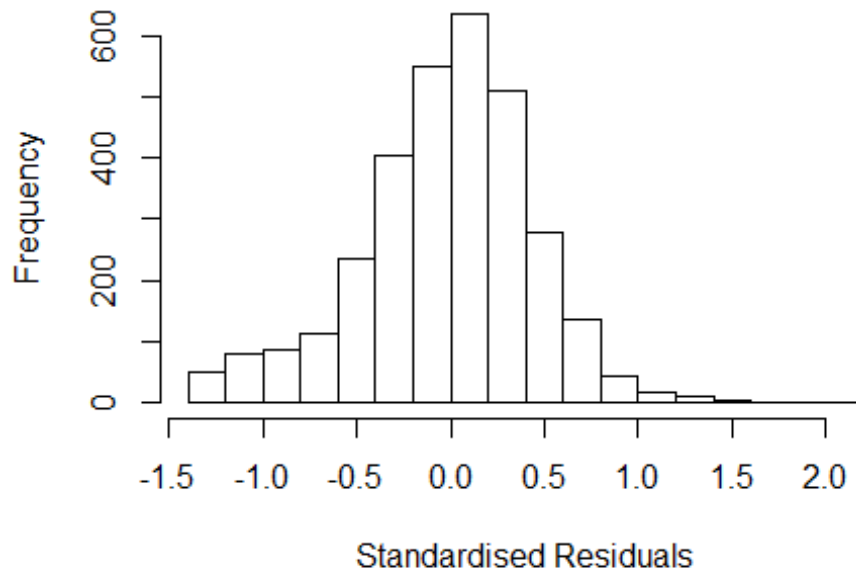
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.285 -0.281  0.018  0.278  2.073
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.28547    0.04327   29.71  <2e-16 ***
## FirstAuthorFemale1  0.02729    0.01601    1.70   0.088 .
## LastAuthorFemale1  0.00181    0.01810    0.10   0.921
## Year1997        -0.01326    0.06876   -0.19   0.847
## Year1998        -0.12187    0.06947   -1.75   0.079 .
## Year1999        -0.08161    0.07044   -1.16   0.247
## Year2000        -0.14289    0.06867   -2.08   0.038 *
## Year2001        -0.08722    0.06435   -1.36   0.175
## Year2002        -0.05707    0.05541   -1.03   0.303
## Year2003        -0.02602    0.05389   -0.48   0.629
## Year2004        -0.07331    0.05074   -1.44   0.149
## Year2005        -0.06356    0.05028   -1.26   0.206
```

```

## Year2006      -0.05015    0.04921   -1.02    0.308
## Year2007      -0.10651    0.04958   -2.15    0.032 *
## Year2008      -0.11294    0.05058   -2.23    0.026 *
## Year2009      -0.11850    0.05055   -2.34    0.019 *
## Year2010      -0.04507    0.04957   -0.91    0.363
## Year2011      -0.10773    0.04922   -2.19    0.029 *
## Year2012      -0.11845    0.05207   -2.27    0.023 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.00827,    Adjusted R-squared:  0.00259
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 597 is an outlier with |weight| = 0 ( < 3.2e-05);
## 280 weights are ~ = 1. The remaining 2878 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0014 0.8570 0.9480 0.8840 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1          1.018
## Year              1.037 16          1.001

```

## Residuals from first author



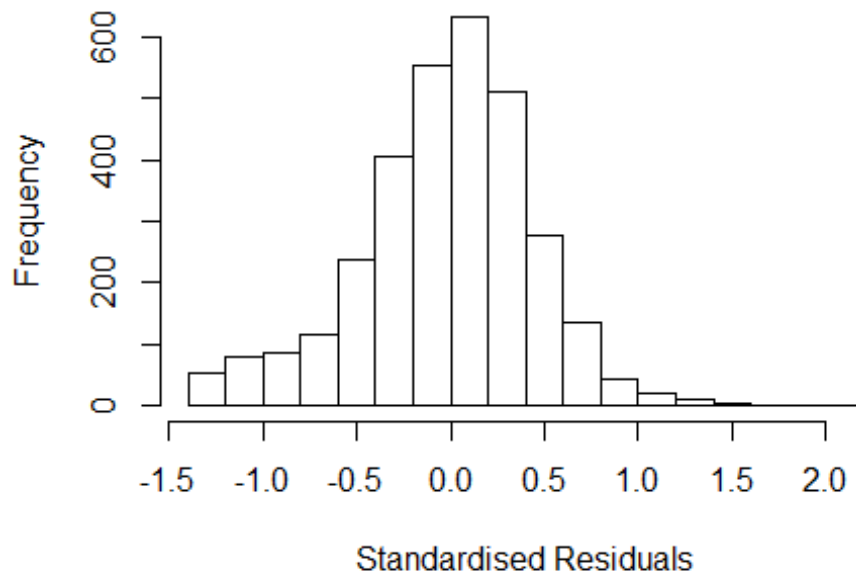
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2859 -0.2805 0.0181 0.2783 2.0731
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2859 0.0430 29.90 <2e-16 ***
## FirstAuthorFemale1 0.0274 0.0161 1.71 0.088 .
## Year1997 -0.0133 0.0687 -0.19 0.847
## Year1998 -0.1218 0.0695 -1.75 0.080 .
## Year1999 -0.0817 0.0704 -1.16 0.246
## Year2000 -0.1430 0.0686 -2.08 0.037 *
## Year2001 -0.0872 0.0643 -1.36 0.175
## Year2002 -0.0572 0.0554 -1.03 0.302
## Year2003 -0.0258 0.0538 -0.48 0.631
## Year2004 -0.0733 0.0507 -1.45 0.148
## Year2005 -0.0635 0.0503 -1.26 0.207
## Year2006 -0.0501 0.0492 -1.02 0.309
```

```

## Year2007          -0.1064      0.0496   -2.15    0.032 *
## Year2008          -0.1130      0.0505   -2.24    0.025 *
## Year2009          -0.1185      0.0506   -2.34    0.019 *
## Year2010          -0.0450      0.0496   -0.91    0.364
## Year2011          -0.1077      0.0492   -2.19    0.029 *
## Year2012          -0.1184      0.0521   -2.27    0.023 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.00827,    Adjusted R-squared:  0.0029
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## observation 597 is an outlier with |weight| = 0 ( < 3.2e-05);
## 283 weights are ~= 1. The remaining 2875 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0013 0.8570 0.9480 0.8840 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.045 1          1.022
## Year          1.045 16          1.001

```

## Residuals from last author



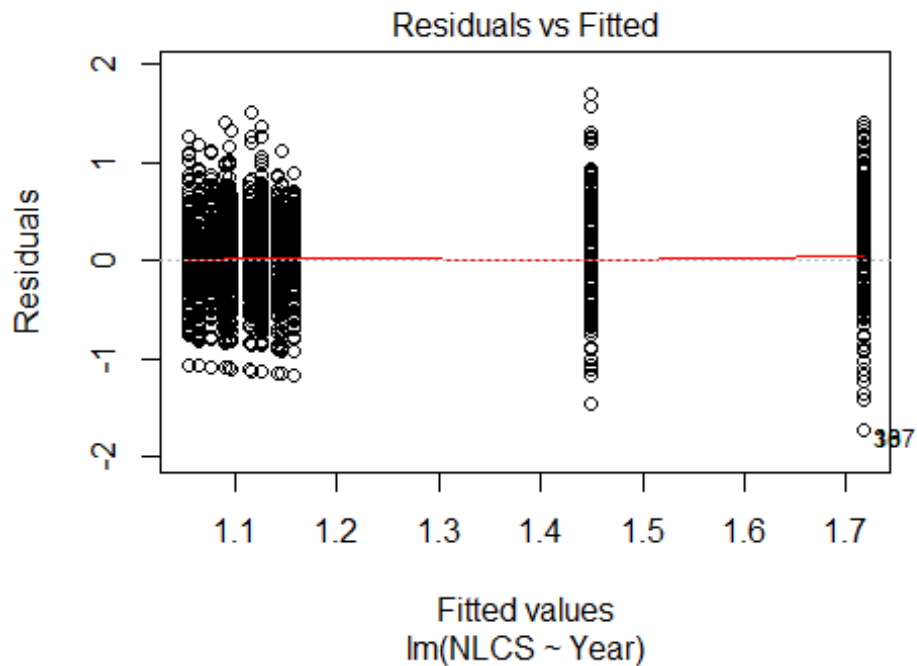
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2951 -0.2765 0.0175 0.2776 2.0656
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29509 0.04309 30.06 <2e-16 ***
## LastAuthorFemale1 0.00498 0.01820 0.27 0.784
## Year1997 -0.01318 0.06882 -0.19 0.848
## Year1998 -0.12039 0.07000 -1.72 0.086 .
## Year1999 -0.07862 0.07030 -1.12 0.263
## Year2000 -0.14465 0.06881 -2.10 0.036 *
## Year2001 -0.08609 0.06447 -1.34 0.182
## Year2002 -0.05491 0.05550 -0.99 0.323
## Year2003 -0.02555 0.05408 -0.47 0.637
## Year2004 -0.07239 0.05088 -1.42 0.155
## Year2005 -0.06159 0.05041 -1.22 0.222
## Year2006 -0.05089 0.04937 -1.03 0.303
```

```

## Year2007          -0.10490      0.04973    -2.11      0.035 *
## Year2008          -0.11318      0.05073    -2.23      0.026 *
## Year2009          -0.11938      0.05080    -2.35      0.019 *
## Year2010          -0.04271      0.04974    -0.86      0.391
## Year2011          -0.10622      0.04939    -2.15      0.032 *
## Year2012          -0.11409      0.05206    -2.19      0.028 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.405
## Multiple R-squared:  0.00736,    Adjusted R-squared:  0.00199
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 597 is an outlier with |weight| = 0 ( < 3.2e-05);
## 289 weights are ~= 1. The remaining 2869 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0023 0.8560 0.9490 0.8830 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.17e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3159"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2809"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 422 376 281 271 279 282 267 229 275 257 302 291 263 305 364
## 2011 2012
## 346 316
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 153 149 144 165 137 104 161 143 135 142 192 188 161 198 262

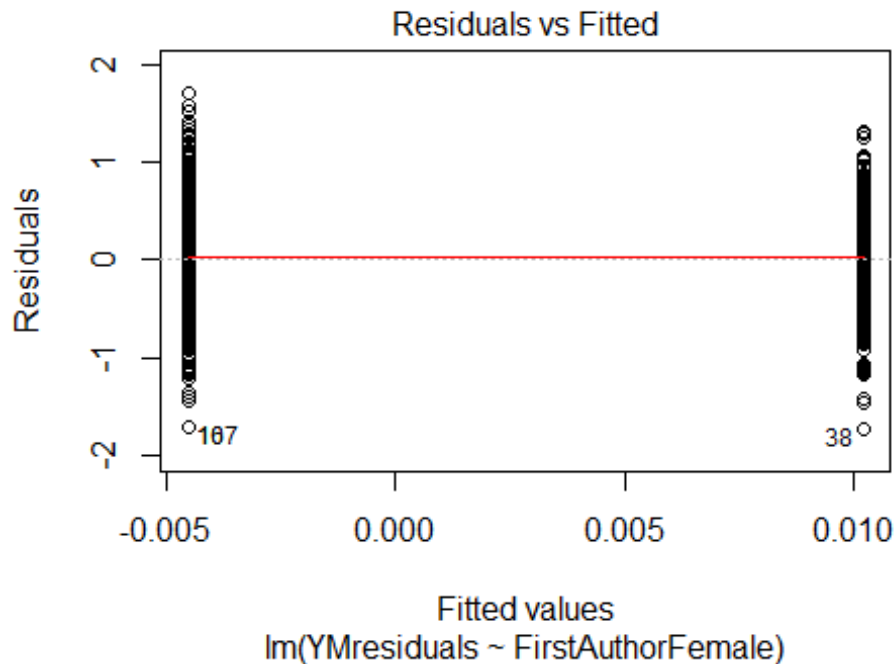
```

```
## 2011 2012
## 237 233
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 142 133 132 145 128 81 122 122 113 114 153 161 139 175 228
## 2011 2012
## 202 192
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 190, df = 16, p-value <2e-16
```



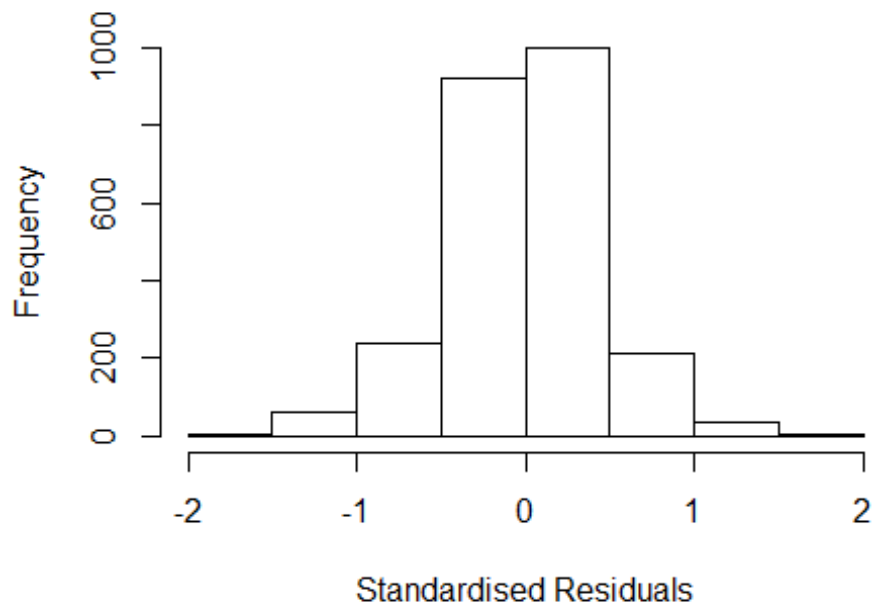
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 3e-04
```





```
## [1] "Female first author team size 2018 geometric mean: 4.43450211830263"
## [1] "Male first author team size 2018 geometric mean: 4.12358485391899"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.96400506393965"
## [1] "Male last author team size 2018 geometric mean: 4.33748648884949"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## LastAuthorFemale  1.051 1      1.025
## UniqueAuthors    1.217 4      1.025
## Year              1.217 16     1.006
```

## Residuals from first and last author and team size



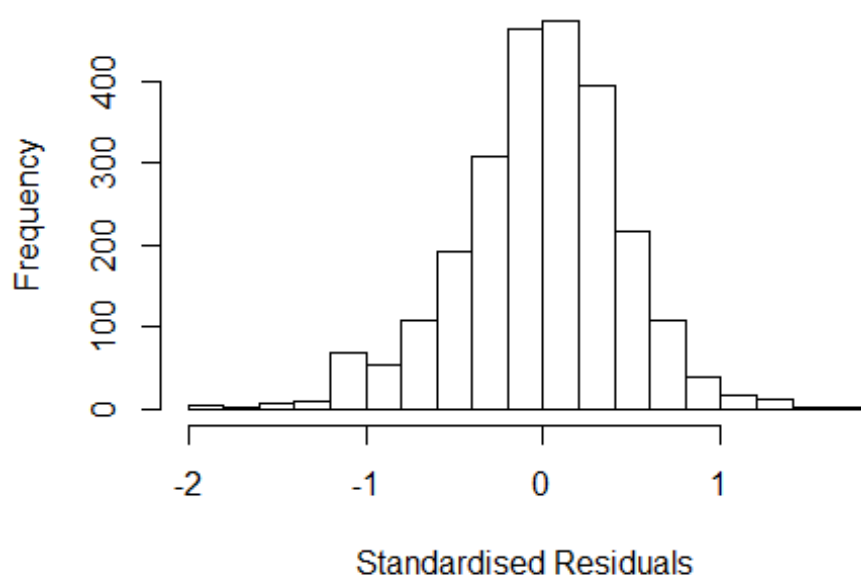
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.96249 -0.26773 0.00497 0.27363 1.81324
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.6505 0.0923 17.88 < 2e-16 ***
## FirstAuthorFemale1 0.0114 0.0182 0.62 0.53239
## LastAuthorFemale1 -0.0465 0.0212 -2.20 0.02810 *
## UniqueAuthors2 0.1325 0.0354 3.74 0.00019 ***
## UniqueAuthors3 0.1934 0.0360 5.38 8.2e-08 ***
## UniqueAuthors4 0.2523 0.0385 6.55 7.1e-11 ***
## UniqueAuthors5 0.3120 0.0358 8.73 < 2e-16 ***
## Year1997 -0.3207 0.1007 -3.18 0.00147 **
## Year1998 -0.7029 0.0927 -7.58 4.7e-14 ***
## Year1999 -0.6495 0.0912 -7.12 1.4e-12 ***
```

```

## Year2000      -0.6554      0.0932      -7.03      2.7e-12 ***
## Year2001      -0.6636      0.0957      -6.93      5.3e-12 ***
## Year2002      -0.6150      0.0952      -6.46      1.3e-10 ***
## Year2003      -0.6750      0.0934      -7.23      6.6e-13 ***
## Year2004      -0.7168      0.0931      -7.70      1.9e-14 ***
## Year2005      -0.7195      0.0955      -7.53      7.0e-14 ***
## Year2006      -0.7381      0.0930      -7.94      3.1e-15 ***
## Year2007      -0.7448      0.0910      -8.18      4.5e-16 ***
## Year2008      -0.7642      0.0918      -8.32      < 2e-16 ***
## Year2009      -0.6946      0.0927      -7.49      9.4e-14 ***
## Year2010      -0.6852      0.0905      -7.57      5.3e-14 ***
## Year2011      -0.8057      0.0905      -8.90      < 2e-16 ***
## Year2012      -0.8051      0.0914      -8.81      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.176, Adjusted R-squared:  0.169
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## observation 26 is an outlier with |weight| = 0 ( < 4e-05);
## 220 weights are ~= 1. The remaining 2261 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0023 0.8660 0.9500 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.03e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.025 1          1.013
## LastAuthorFemale 1.027 1          1.013
## Year      1.046 16          1.001

```

## Residuals from first and last author



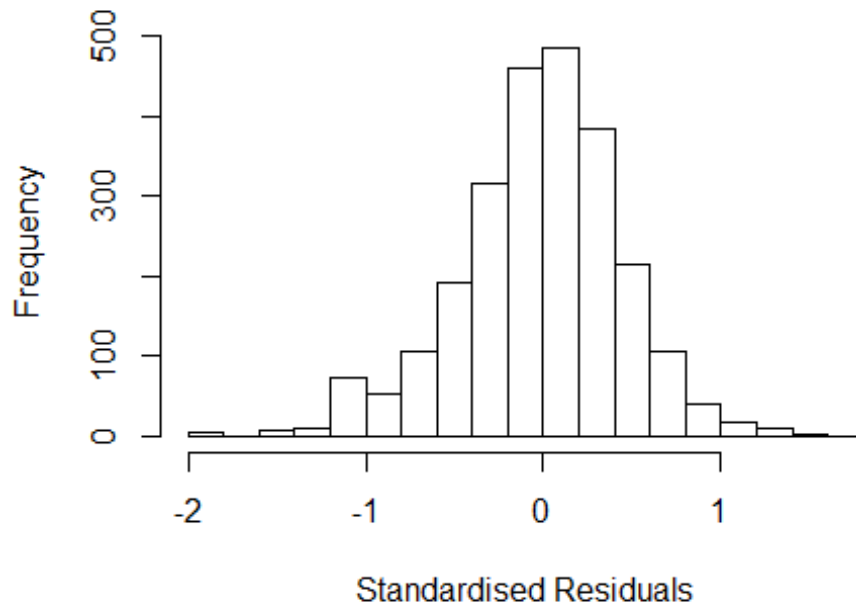
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.80849 -0.27051 0.00729 0.27765 1.65479
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8085 0.0908 19.92 < 2e-16 ***
## FirstAuthorFemale1 0.0336 0.0184 1.82 0.0687 .
## LastAuthorFemale1 -0.0560 0.0216 -2.59 0.0096 **
## Year1997 -0.3203 0.1058 -3.03 0.0025 **
## Year1998 -0.7091 0.0966 -7.34 2.8e-13 ***
## Year1999 -0.6471 0.0957 -6.76 1.7e-11 ***
## Year2000 -0.6440 0.0982 -6.56 6.6e-11 ***
## Year2001 -0.6534 0.0983 -6.65 3.7e-11 ***
## Year2002 -0.5880 0.0998 -5.89 4.3e-09 ***
## Year2003 -0.6474 0.0976 -6.63 4.0e-11 ***
## Year2004 -0.7082 0.0978 -7.24 6.1e-13 ***
## Year2005 -0.6838 0.0999 -6.84 9.8e-12 ***
```

```

## Year2006          -0.7013      0.0972    -7.21  7.2e-13 ***
## Year2007          -0.7154      0.0951    -7.52  7.6e-14 ***
## Year2008          -0.7322      0.0961    -7.62  3.7e-14 ***
## Year2009          -0.6540      0.0969    -6.75  1.8e-11 ***
## Year2010          -0.6557      0.0948    -6.92  5.9e-12 ***
## Year2011          -0.7613      0.0949    -8.02  1.6e-15 ***
## Year2012          -0.7638      0.0959    -7.96  2.5e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.139, Adjusted R-squared:  0.132
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 224 weights are ~= 1. The remaining 2258 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0132 0.8600 0.9490 0.8870 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001

```

## Residuals from first author



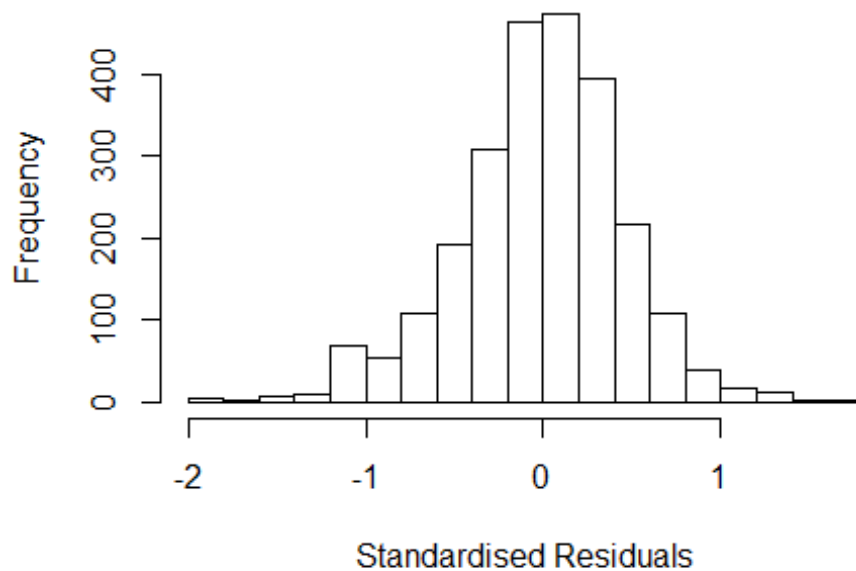
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82878 -0.27334  0.00814  0.27649  1.66076
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.8025     0.0912   19.77 < 2e-16 ***
## FirstAuthorFemale1  0.0262     0.0186    1.41  0.1577
## Year1997         -0.3203     0.1061   -3.02  0.0026 **
## Year1998         -0.7114     0.0970   -7.33 3.1e-13 ***
## Year1999         -0.6471     0.0961   -6.74 2.0e-11 ***
## Year2000         -0.6440     0.0987   -6.53 8.2e-11 ***
## Year2001         -0.6567     0.0989   -6.64 3.9e-11 ***
## Year2002         -0.5941     0.1003   -5.92 3.6e-09 ***
## Year2003         -0.6502     0.0981   -6.63 4.2e-11 ***
## Year2004         -0.7085     0.0983   -7.20 7.7e-13 ***
## Year2005         -0.6856     0.1004   -6.83 1.1e-11 ***
## Year2006         -0.7045     0.0976   -7.21 7.2e-13 ***
```

```

## Year2007          -0.7219      0.0956   -7.55  6.1e-14 ***
## Year2008          -0.7369      0.0966   -7.63  3.4e-14 ***
## Year2009          -0.6590      0.0975   -6.76  1.7e-11 ***
## Year2010          -0.6588      0.0953   -6.92  5.9e-12 ***
## Year2011          -0.7649      0.0954   -8.02  1.6e-15 ***
## Year2012          -0.7691      0.0963   -7.98  2.2e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.136, Adjusted R-squared:  0.13
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 235 weights are ~= 1. The remaining 2247 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0089 0.8600 0.9480 0.8870 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.022 1          1.011
## Year              1.022 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.81536 -0.27536 0.00707 0.27717 1.64497
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8154 0.0904 20.08 < 2e-16 ***
## LastAuthorFemale1 -0.0500 0.0216 -2.31 0.0209 *
## Year1997 -0.3173 0.1057 -3.00 0.0027 **
## Year1998 -0.7045 0.0965 -7.30 3.8e-13 ***
## Year1999 -0.6456 0.0955 -6.76 1.7e-11 ***
## Year2000 -0.6444 0.0981 -6.57 6.2e-11 ***
## Year2001 -0.6537 0.0982 -6.66 3.4e-11 ***
## Year2002 -0.5855 0.0997 -5.87 4.9e-09 ***
## Year2003 -0.6440 0.0975 -6.60 5.0e-11 ***
## Year2004 -0.7065 0.0978 -7.22 6.7e-13 ***
## Year2005 -0.6798 0.0999 -6.81 1.2e-11 ***
## Year2006 -0.6999 0.0970 -7.22 7.0e-13 ***
```

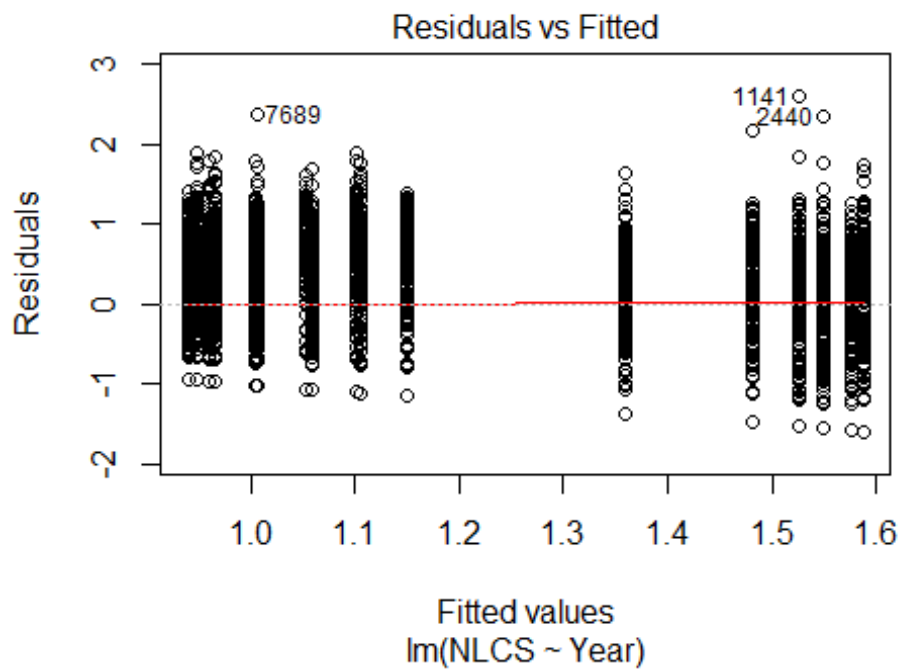


```

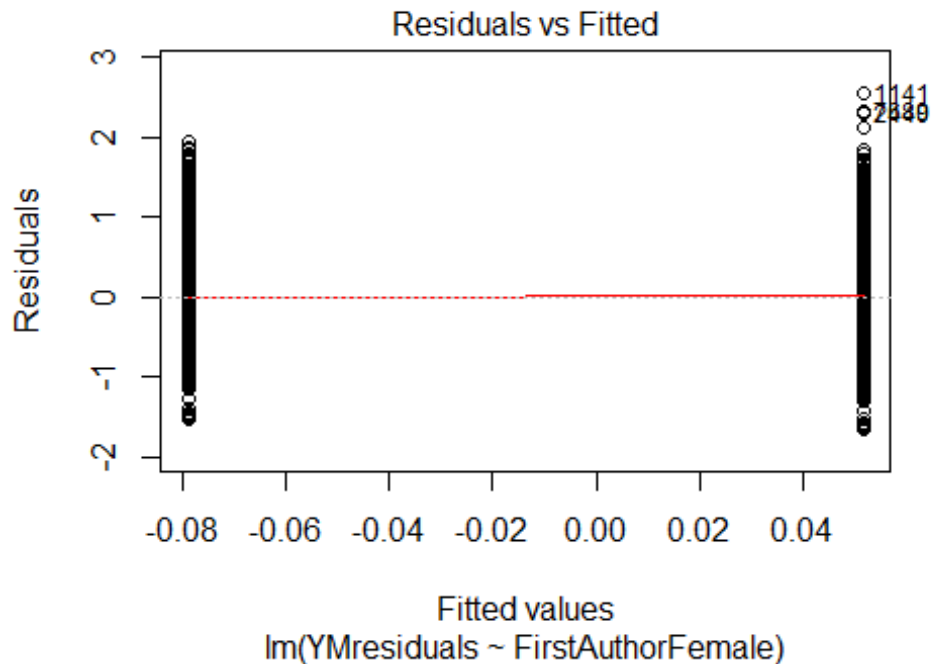
## Year2007          -0.7120      0.0950   -7.50  9.1e-14 ***
## Year2008          -0.7292      0.0960   -7.60  4.3e-14 ***
## Year2009          -0.6523      0.0967   -6.74  1.9e-11 ***
## Year2010          -0.6515      0.0947   -6.88  7.5e-12 ***
## Year2011          -0.7566      0.0947   -7.99  2.1e-15 ***
## Year2012          -0.7620      0.0958   -7.95  2.7e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.138, Adjusted R-squared:  0.132
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 228 weights are ~= 1. The remaining 2254 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0126 0.8620 0.9490 0.8880 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.03e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2482"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2900"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 435 593 611 613 657 744 581 500 446 495 557 668 579 407 390
## 2011 2012
## 403 452
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 281 288 264 269 349 349 440 401 371 399 453 525 466 356 336
## 2011 2012

```

```
## 359 388
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 274 281 256 254 332 319 409 367 339 362 408 487 420 339 319
## 2011 2012
## 340 372
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 63, df = 16, p-value = 1e-07
```

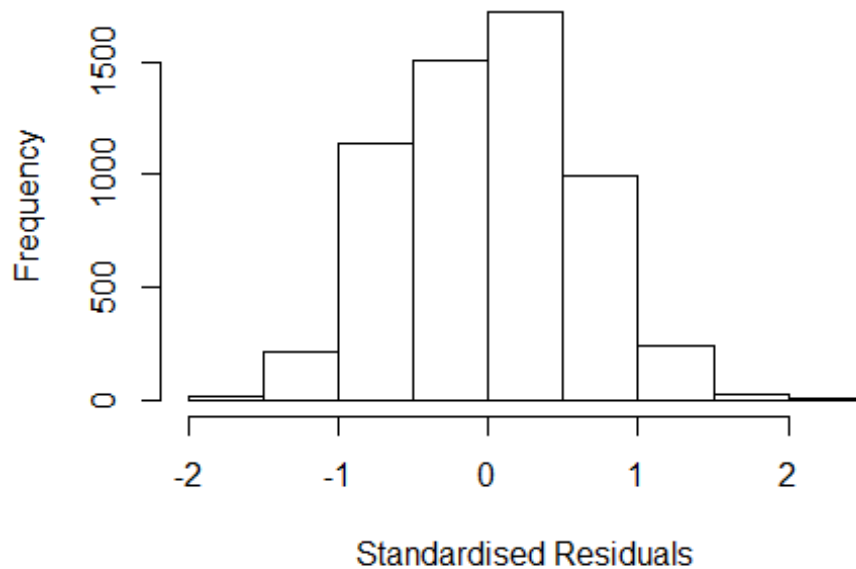


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.34659451634759"
## [1] "Male first author team size 2018 geometric mean: 1.97924959897119"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9500, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.27323699858191"
## [1] "Male last author team size 2018 geometric mean: 2.13258298618372"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.430 1          1.196
## LastAuthorFemale  1.403 1          1.184
## UniqueAuthors    1.154 4          1.018
## Year              1.174 16         1.005
```

## Residuals from first and last author and team size



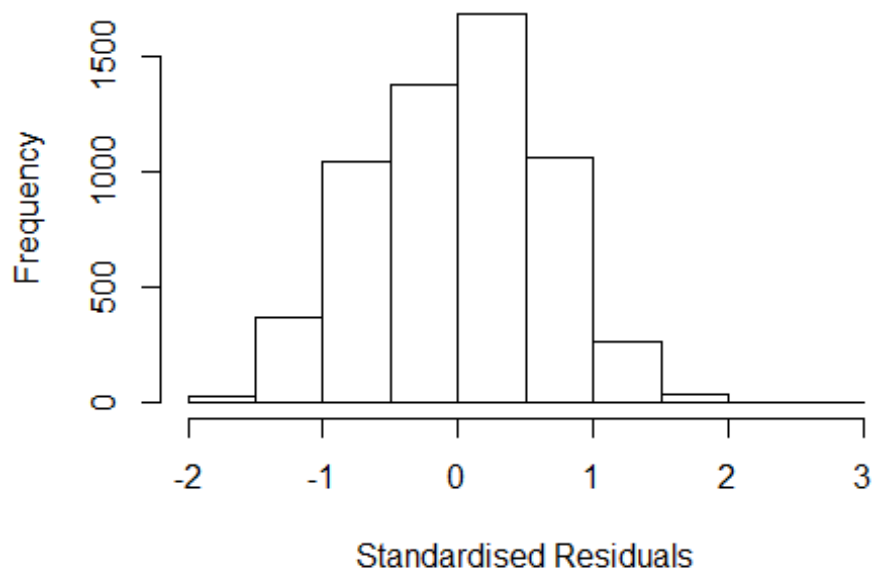
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.9350 -0.4507 0.0153 0.4367 2.2510
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4144 0.0454 31.13 < 2e-16 ***
## FirstAuthorFemale1 0.1016 0.0205 4.96 7.3e-07 ***
## LastAuthorFemale1 0.0160 0.0199 0.80 0.421
## UniqueAuthors2 0.3219 0.0214 15.03 < 2e-16 ***
## UniqueAuthors3 0.4831 0.0260 18.59 < 2e-16 ***
## UniqueAuthors4 0.5045 0.0323 15.64 < 2e-16 ***
## UniqueAuthors5 0.5848 0.0312 18.72 < 2e-16 ***
## Year1997 -0.0910 0.0565 -1.61 0.107
## Year1998 -0.0801 0.0607 -1.32 0.187
## Year1999 -0.0341 0.0563 -0.60 0.545
```

```

## Year2000          -0.1602      0.0556   -2.88    0.004 **
## Year2001          -0.3292      0.0558   -5.90   3.8e-09 ***
## Year2002          -0.7003      0.0551  -12.71   < 2e-16 ***
## Year2003          -0.7837      0.0552  -14.20   < 2e-16 ***
## Year2004          -0.7862      0.0545  -14.42   < 2e-16 ***
## Year2005          -0.7955      0.0547  -14.54   < 2e-16 ***
## Year2006          -0.7730      0.0538  -14.38   < 2e-16 ***
## Year2007          -0.7015      0.0526  -13.34   < 2e-16 ***
## Year2008          -0.7270      0.0561  -12.96   < 2e-16 ***
## Year2009          -0.5987      0.0581  -10.30   < 2e-16 ***
## Year2010          -0.5888      0.0564  -10.45   < 2e-16 ***
## Year2011          -0.6421      0.0582  -11.03   < 2e-16 ***
## Year2012          -0.6744      0.0570  -11.83   < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.647
## Multiple R-squared:  0.223, Adjusted R-squared:  0.22
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 456 weights are ~= 1. The remaining 5422 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.201  0.879  0.949  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.451 1      1.205
## LastAuthorFemale  1.437 1      1.199
## Year              1.029 16      1.001

```

## Residuals from first and last author



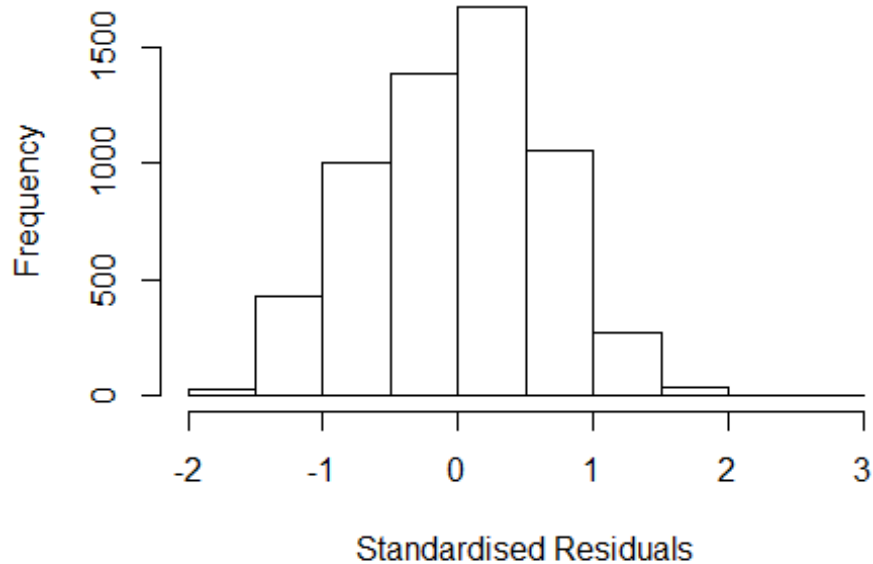
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1141 0031664637 4.128 1998      2739      3      2.522
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.680 -0.490  0.034  0.469  2.522
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.52548    0.04434   34.41 < 2e-16 ***
## FirstAuthorFemale1 0.15417    0.02221    6.94 4.3e-12 ***
## LastAuthorFemale1 -0.02840    0.02158   -1.32 0.18808
## Year1997        -0.09551    0.05489   -1.74 0.08190 .
## Year1998        -0.07315    0.05877   -1.24 0.21335
## Year1999        -0.00242    0.05541   -0.04 0.96517
## Year2000        -0.05178    0.05481   -0.94 0.34490
## Year2001        -0.19834    0.05408   -3.67 0.00025 ***
## Year2002        -0.60756    0.05471  -11.11 < 2e-16 ***
## Year2003        -0.65611    0.05590  -11.74 < 2e-16 ***
## Year2004        -0.68072    0.05685  -11.97 < 2e-16 ***
## Year2005        -0.67860    0.05530  -12.27 < 2e-16 ***
```

```

## Year2006      -0.65185    0.05498   -11.86 < 2e-16 ***
## Year2007      -0.55291    0.05231   -10.57 < 2e-16 ***
## Year2008      -0.61994    0.05757   -10.77 < 2e-16 ***
## Year2009      -0.49790    0.06039    -8.24 < 2e-16 ***
## Year2010      -0.48322    0.05751    -8.40 < 2e-16 ***
## Year2011      -0.54145    0.05971    -9.07 < 2e-16 ***
## Year2012      -0.55876    0.06045    -9.24 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.689
## Multiple R-squared:  0.126, Adjusted R-squared:  0.123
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 484 weights are ~= 1. The remaining 5394 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.151  0.868  0.948  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.02 1         1.010
## Year              1.02 16         1.001

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1141 0031664637 4.128 1998      2739      3      2.522
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6560 -0.4914  0.0362  0.4671  2.5442
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.51880    0.04400   34.52 < 2e-16 ***
## FirstAuthorFemale1 0.13723    0.01861    7.38 1.9e-13 ***
## Year1997      -0.09424    0.05487   -1.72 0.08597 .
## Year1998      -0.07221    0.05871   -1.23 0.21876
## Year1999      -0.00112    0.05540   -0.02 0.98383
## Year2000      -0.05133    0.05483   -0.94 0.34929
## Year2001      -0.19829    0.05409   -3.67 0.00025 ***
## Year2002      -0.60684    0.05472  -11.09 < 2e-16 ***
## Year2003      -0.65573    0.05590  -11.73 < 2e-16 ***
## Year2004      -0.67941    0.05682  -11.96 < 2e-16 ***
## Year2005      -0.67723    0.05525  -12.26 < 2e-16 ***
## Year2006      -0.65049    0.05499  -11.83 < 2e-16 ***
```



```

## Year2007          -0.55177      0.05231  -10.55  < 2e-16 ***
## Year2008          -0.61960      0.05759  -10.76  < 2e-16 ***
## Year2009          -0.49693      0.06043   -8.22  2.4e-16 ***
## Year2010          -0.48148      0.05754   -8.37  < 2e-16 ***
## Year2011          -0.53982      0.05974   -9.04  < 2e-16 ***
## Year2012          -0.55833      0.06051   -9.23  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.688
## Multiple R-squared:  0.126, Adjusted R-squared:  0.123
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 489 weights are ~= 1. The remaining 5389 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.143  0.868  0.948  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year            1.012 16          1.000

## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1141 0031664637 4.128 1998      2739      3      2.522
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.6223 -0.4820  0.0347  0.4750  2.6311

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.55918    0.04432   35.18 < 2e-16 ***
## LastAuthorFemale1 0.06004    0.01831    3.28 0.00105 **
## Year1997       -0.08993    0.05505   -1.63 0.10241
## Year1998       -0.06225    0.05904   -1.05 0.29181
## Year1999        0.00307    0.05535    0.06 0.95583
## Year2000       -0.04226    0.05509   -0.77 0.44310
## Year2001       -0.19719    0.05423   -3.64 0.00028 ***
## Year2002       -0.60343    0.05486  -11.00 < 2e-16 ***
## Year2003       -0.65155    0.05616  -11.60 < 2e-16 ***
## Year2004       -0.67105    0.05724  -11.72 < 2e-16 ***
## Year2005       -0.67092    0.05551  -12.09 < 2e-16 ***
## Year2006       -0.64424    0.05542  -11.62 < 2e-16 ***
## Year2007       -0.53986    0.05258  -10.27 < 2e-16 ***
## Year2008       -0.60441    0.05755  -10.50 < 2e-16 ***
## Year2009       -0.48470    0.06077   -7.98 1.8e-15 ***
## Year2010       -0.46053    0.05809   -7.93 2.6e-15 ***
## Year2011       -0.52289    0.06030   -8.67 < 2e-16 ***
## Year2012       -0.54415    0.06087   -8.94 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.692
## Multiple R-squared:  0.118, Adjusted R-squared:  0.116
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 460 weights are ~= 1. The remaining 5418 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.116  0.866  0.948  0.917  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.70e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0        1000         0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5878"
## [1] ""

```

```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2901"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2   34   35   41   31   30   34   29   18   20   42   34   53   41   44
## 2011 2012
##   44   43
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    1    1    0    1    0    2    8    2    8   13   10   32   27   19
## 2011 2012
##   22   17
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    1    1    0    1    0    2    8    1    7   13   10   27   20   15
## 2011 2012
##   19   15
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.88902194973031"
## [1] "Male first author team size 2018 geometric mean: 3.01345262384007"

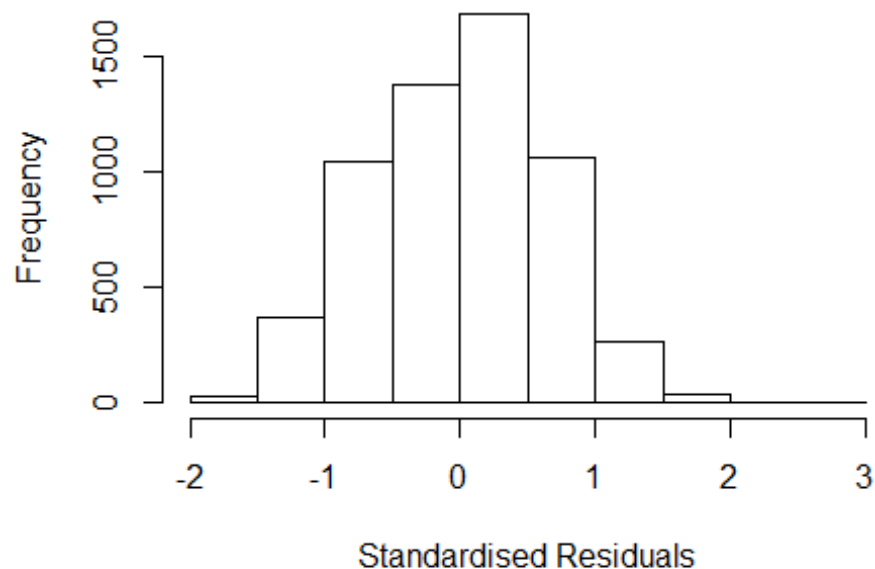
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.8513481795695"
## [1] "Male last author team size 2018 geometric mean: 3.05326821786485"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

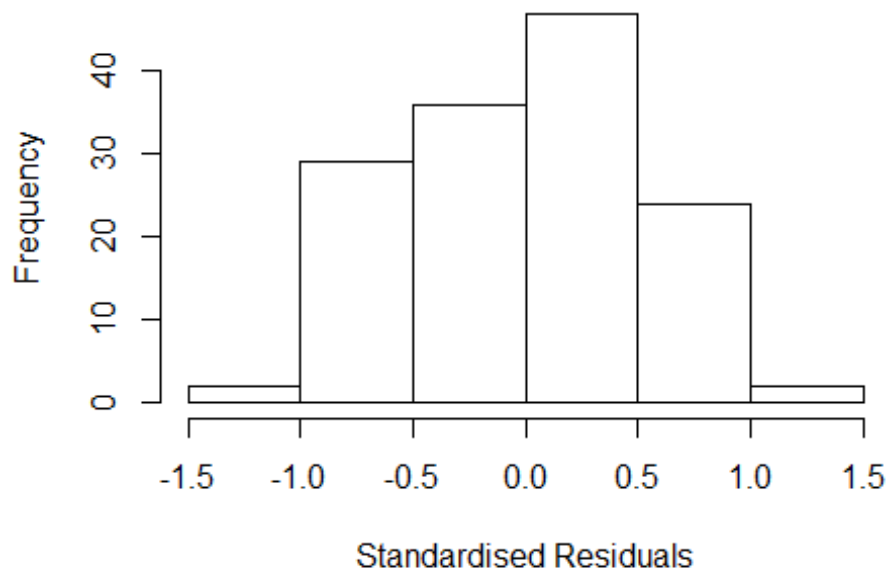
```

## Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -6.785e+15  1      NaN
## LastAuthorFemale  -5.772e+15  1      NaN
## UniqueAuthors      1.472e+01  4      1.399
## Year              -1.159e+17 13      NaN
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1841 -0.3657 0.0137 0.3408 1.4624
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.56e+00 5.54e-08 2.81e+07 < 2e-16 ***
## FirstAuthorFemale1 2.45e-01 1.28e-01 1.92e+00 0.057 .
## LastAuthorFemale1 -1.18e-01 1.18e-01 -1.00e+00 0.317
## UniqueAuthors2 -5.60e-02 1.57e-01 -3.60e-01 0.721
## UniqueAuthors3 3.45e-01 1.66e-01 2.08e+00 0.040 *
## UniqueAuthors4 1.34e-01 1.67e-01 8.00e-01 0.423
## UniqueAuthors5 2.99e-01 1.48e-01 2.03e+00 0.045 *
## Year1998 -1.69e+00 1.45e-01 -1.16e+01 < 2e-16 ***
## Year2000 -1.56e+00 2.88e-15 -5.42e+14 < 2e-16 ***
## Year2002 -6.77e-01 2.74e-01 -2.47e+00 0.015 *
```

```

## Year2003      -1.05e+00  2.40e-01 -4.38e+00  2.5e-05 ***
## Year2004      -1.21e+00  4.02e-08 -3.00e+07  < 2e-16 ***
## Year2005      -1.19e+00  1.91e-01 -6.23e+00  7.2e-09 ***
## Year2006      -1.04e+00  2.05e-01 -5.08e+00  1.4e-06 ***
## Year2007      -1.08e+00  2.13e-01 -5.08e+00  1.4e-06 ***
## Year2008      -9.98e-01  2.22e-01 -4.49e+00  1.6e-05 ***
## Year2009      -9.38e-01  1.51e-01 -6.23e+00  7.1e-09 ***
## Year2010      -9.40e-01  1.81e-01 -5.20e+00  8.5e-07 ***
## Year2011      -8.47e-01  1.68e-01 -5.05e+00  1.6e-06 ***
## Year2012      -9.14e-01  1.99e-01 -4.60e+00  1.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.183, Adjusted R-squared:  0.0533
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 122 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.474  0.884  0.940  0.916  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.14e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.028e+15  1      3.206e+07
## LastAuthorFemale  9.088e+14  1      3.015e+07
## Year      1.660e+15 13      3.849e+00
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =

```

```

AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q   Median      3Q      Max
## -9.84e-01 -3.93e-01 -5.55e-16  4.25e-01  1.38e+00
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.56e+00   7.63e-08  2.04e+07 < 2e-16 ***
## FirstAuthorFemale1  3.39e-01   1.30e-01  2.60e+00  0.011 *
## LastAuthorFemale1 -1.63e-01   1.23e-01 -1.33e+00  0.187
## Year1998         -1.74e+00   1.38e-01 -1.26e+01 < 2e-16 ***
## Year2000         -1.56e+00   5.14e-08 -3.03e+07 < 2e-16 ***
## Year2002         -6.77e-01   2.72e-01 -2.49e+00  0.014 *
## Year2003         -1.09e+00   2.39e-01 -4.57e+00  1.2e-05 ***
## Year2004         -1.21e+00   6.41e-08 -1.88e+07 < 2e-16 ***
## Year2005         -1.04e+00   1.68e-01 -6.16e+00  9.4e-09 ***
## Year2006         -1.02e+00   1.89e-01 -5.37e+00  3.7e-07 ***
## Year2007         -9.67e-01   2.28e-01 -4.25e+00  4.2e-05 ***
## Year2008         -9.14e-01   1.92e-01 -4.75e+00  5.4e-06 ***
## Year2009         -8.15e-01   1.25e-01 -6.49e+00  1.8e-09 ***
## Year2010         -8.52e-01   1.54e-01 -5.54e+00  1.7e-07 ***
## Year2011         -8.13e-01   1.45e-01 -5.60e+00  1.3e-07 ***
## Year2012         -8.06e-01   1.71e-01 -4.72e+00  6.2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.596
## Multiple R-squared:  0.122, Adjusted R-squared:  0.016
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 122 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.572  0.877   0.947   0.921   0.981   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.14e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats

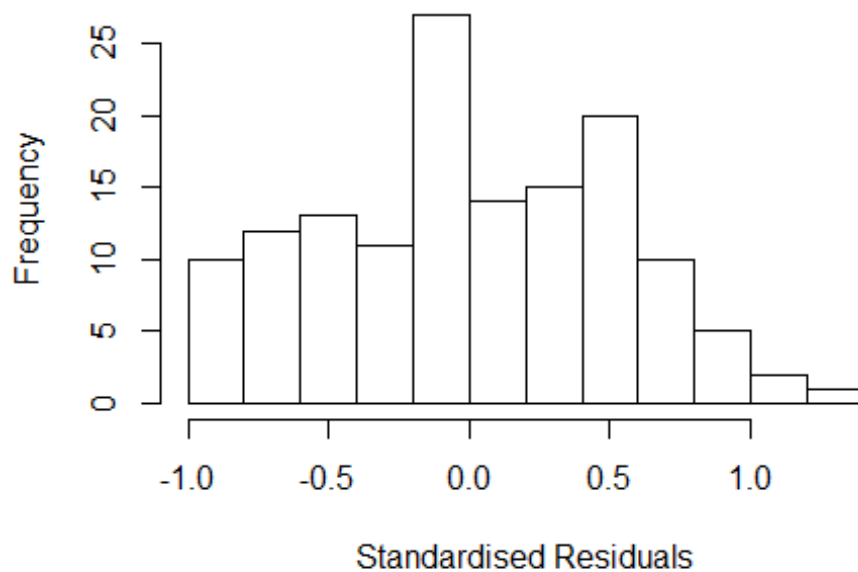
```

```
## "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in cov2cor(v): diag(.) had 0 or NA entries; non-finite result is
## doubtful
```

### Residuals from first and last author



```
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1          NaN
## Year              NaN 13         NaN

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields  residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97161 -0.39521  0.00195  0.40438  1.43217
##
```



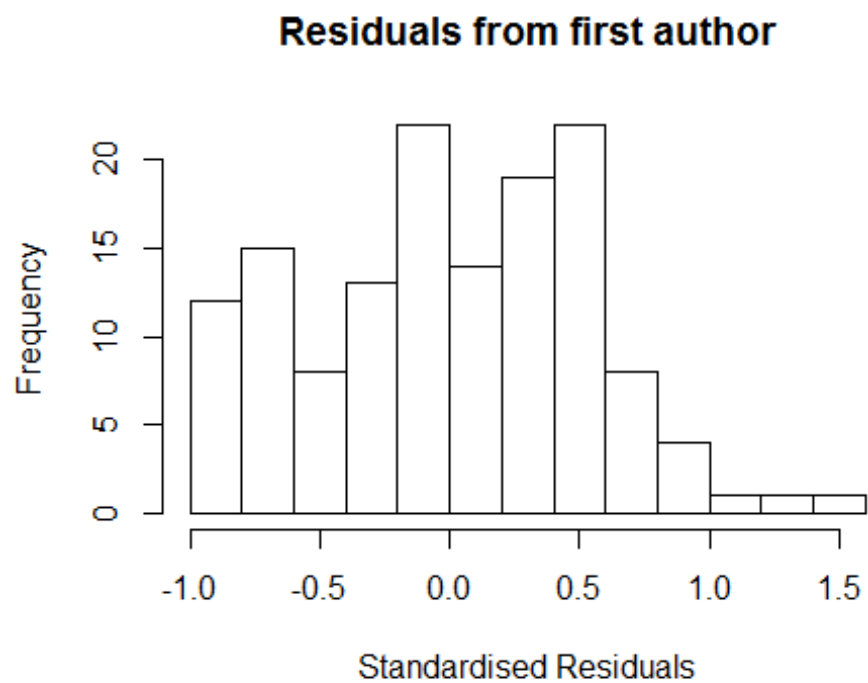
```

## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.560      0.000    Inf < 2e-16 ***
## FirstAuthorFemale1 0.262      0.119     2.19 0.03002 *
## Year1998         -1.822      0.119   -15.25 < 2e-16 ***
## Year2000         -1.560      0.000   -Inf < 2e-16 ***
## Year2002         -0.677      0.272     -2.49 0.01407 *
## Year2003         -1.160      0.237     -4.90 2.9e-06 ***
## Year2004         -1.207      0.000   -Inf < 2e-16 ***
## Year2005         -1.120      0.151     -7.41 1.6e-11 ***
## Year2006         -1.064      0.184     -5.78 5.7e-08 ***
## Year2007         -0.949      0.240     -3.95 0.00013 ***
## Year2008         -0.968      0.187     -5.18 8.6e-07 ***
## Year2009         -0.851      0.130     -6.54 1.4e-09 ***
## Year2010         -0.868      0.156     -5.56 1.5e-07 ***
## Year2011         -0.852      0.143     -5.95 2.6e-08 ***
## Year2012         -0.872      0.148     -5.88 3.4e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0114
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 130 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.537 0.889 0.948 0.924 0.984 0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           7.14e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

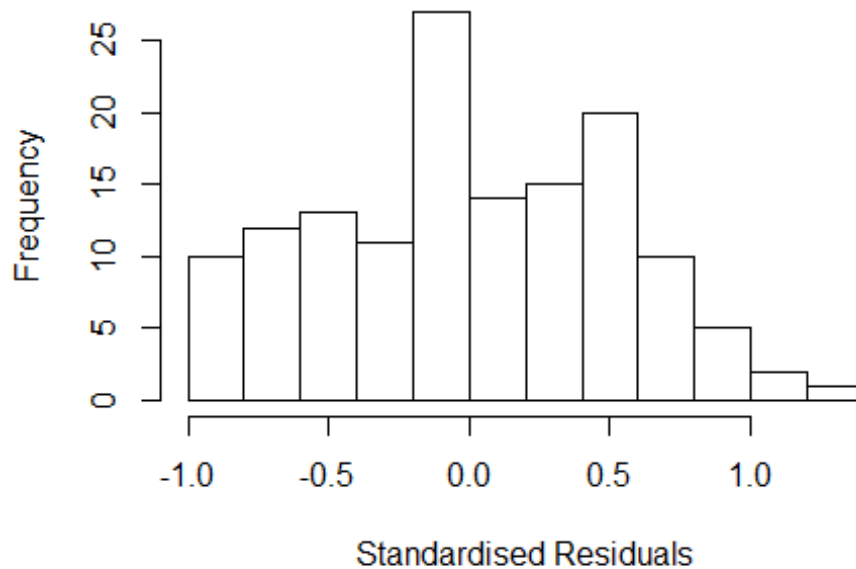
```

```
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale  NaN  1          NaN
## Year              NaN 13          NaN
```

## Residuals from last author



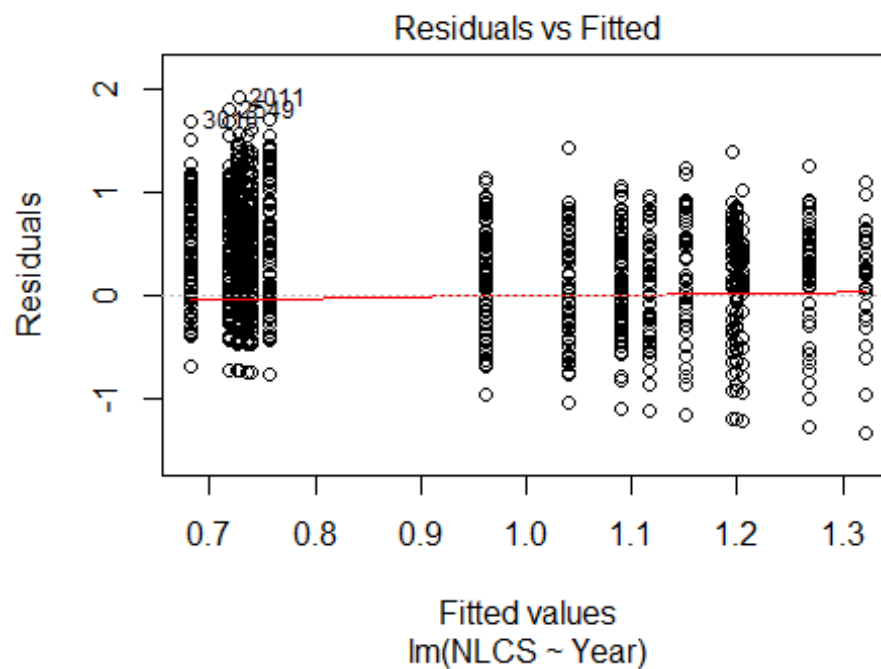
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.89736 -0.48046 -0.00361 0.43726 1.24913
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.56e+00 5.33e-08 2.93e+07 < 2e-16 ***
## LastAuthorFemale1 -2.16e-02 1.14e-01 -1.90e-01 0.8504
## Year1998 -1.54e+00 1.14e-01 -1.34e+01 < 2e-16 ***
## Year2000 -1.56e+00 1.06e-08 -1.47e+08 < 2e-16 ***
## Year2002 -6.77e-01 2.70e-01 -2.51e+00 0.0134 *
## Year2003 -9.33e-01 2.14e-01 -4.35e+00 2.8e-05 ***
## Year2004 -1.21e+00 0.00e+00 -Inf < 2e-16 ***
## Year2005 -1.00e+00 1.75e-01 -5.74e+00 6.9e-08 ***
## Year2006 -8.52e-01 1.94e-01 -4.40e+00 2.3e-05 ***
## Year2007 -7.63e-01 2.43e-01 -3.14e+00 0.0021 **
## Year2008 -7.33e-01 1.64e-01 -4.48e+00 1.7e-05 ***
## Year2009 -6.49e-01 1.22e-01 -5.31e+00 4.9e-07 ***
```

```

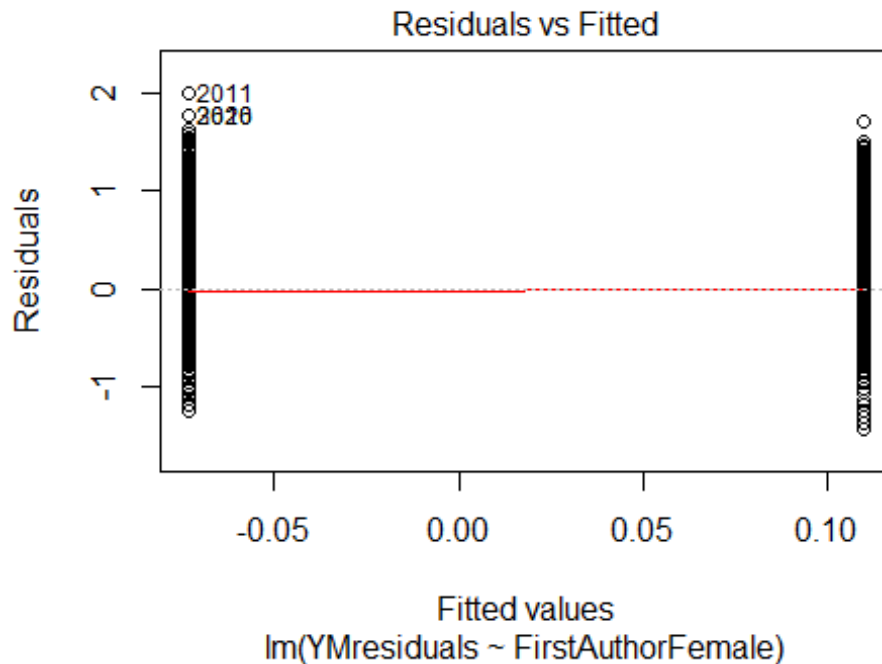
## Year2010          -7.47e-01   1.61e-01 -4.65e+00   8.2e-06 ***
## Year2011          -6.41e-01   1.37e-01 -4.68e+00   7.3e-06 ***
## Year2012          -7.59e-01   1.70e-01 -4.47e+00   1.8e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.0718, Adjusted R-squared:  -0.0321
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 119 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.667  0.865  0.934  0.922  0.977  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.14e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 140"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2902"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   87   96   89   92  127  129  113  119   90  116  195  195  220  284  246
## 2011 2012
##  237  215
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   36   50   51   54   45   48   71   97   69   96  165  151  183  235  203
## 2011 2012
##  193  174
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010

```

```
## 34 43 48 49 42 43 66 80 62 88 152 141 159 218 190
## 2011 2012
## 179 162
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 16, df = 16, p-value = 0.5
```

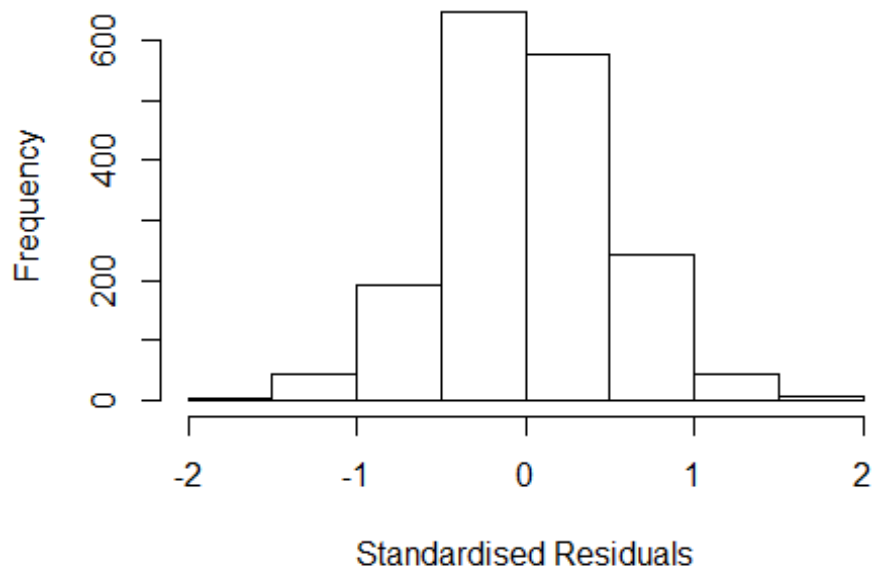


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.4, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.55274531465407"
## [1] "Male first author team size 2018 geometric mean: 3.40225948276024"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.3122404240354"
## [1] "Male last author team size 2018 geometric mean: 3.50920262230201"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.006
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.940 1          1.393
## LastAuthorFemale  2.316 1          1.522
## UniqueAuthors    1.750 4          1.072
## Year             1.464 16          1.012
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5141 -0.3300 -0.0189  0.3580  1.9657
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6936    0.1006     6.90 7.4e-12 ***
## FirstAuthorFemale1 -0.0184    0.0353    -0.52 0.60186
## LastAuthorFemale1 -0.1379    0.0373    -3.70 0.00022 ***
## UniqueAuthors2     0.3474    0.0401     8.66 < 2e-16 ***
## UniqueAuthors3     0.6167    0.0489    12.61 < 2e-16 ***
## UniqueAuthors4     0.8097    0.0544    14.88 < 2e-16 ***
## UniqueAuthors5     0.9598    0.0383    25.03 < 2e-16 ***
## Year1997          0.0722    0.1164     0.62 0.53505
## Year1998          0.0269    0.1167     0.23 0.81755
## Year1999          0.0423    0.1228     0.34 0.73068
```

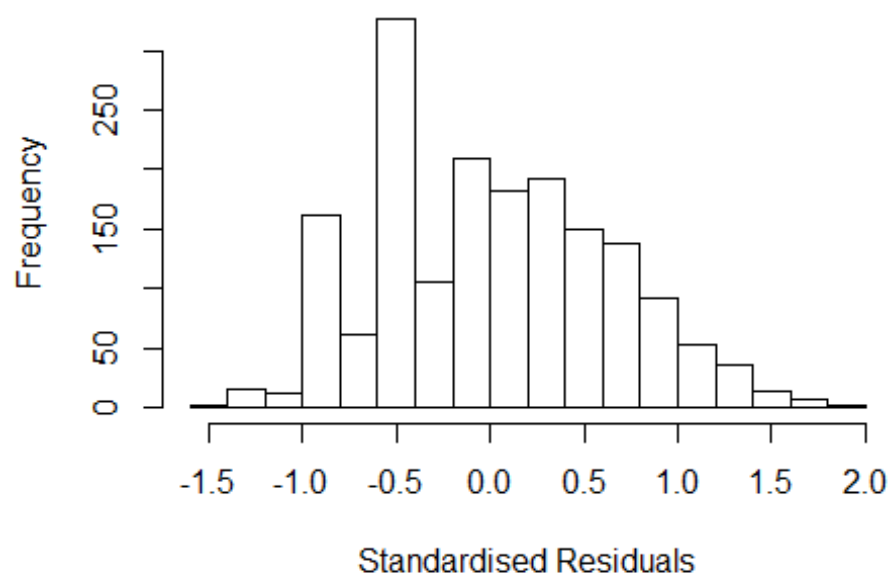
```

## Year2000          -0.0392      0.1136   -0.35   0.72991
## Year2001          -0.0793      0.1184   -0.67   0.50320
## Year2002           0.1314      0.1145    1.15   0.25131
## Year2003          -0.0323      0.1085   -0.30   0.76603
## Year2004          -0.1360      0.1136   -1.20   0.23146
## Year2005          -0.1393      0.1125   -1.24   0.21580
## Year2006          -0.2871      0.1057   -2.72   0.00667 **
## Year2007          -0.3010      0.1044   -2.88   0.00398 **
## Year2008          -0.3013      0.1057   -2.85   0.00443 **
## Year2009          -0.2745      0.1024   -2.68   0.00738 **
## Year2010          -0.2807      0.1051   -2.67   0.00763 **
## Year2011          -0.3186      0.1050   -3.03   0.00245 **
## Year2012          -0.2605      0.1083   -2.41   0.01625 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.463, Adjusted R-squared:  0.456
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 101 weights are ~= 1. The remaining 1655 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0678 0.8740 0.9470 0.9020 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.69e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.789 1 1.338
## LastAuthorFemale 1.933 1 1.390
## Year 1.225 16 1.006

```



## Residuals from first and last author



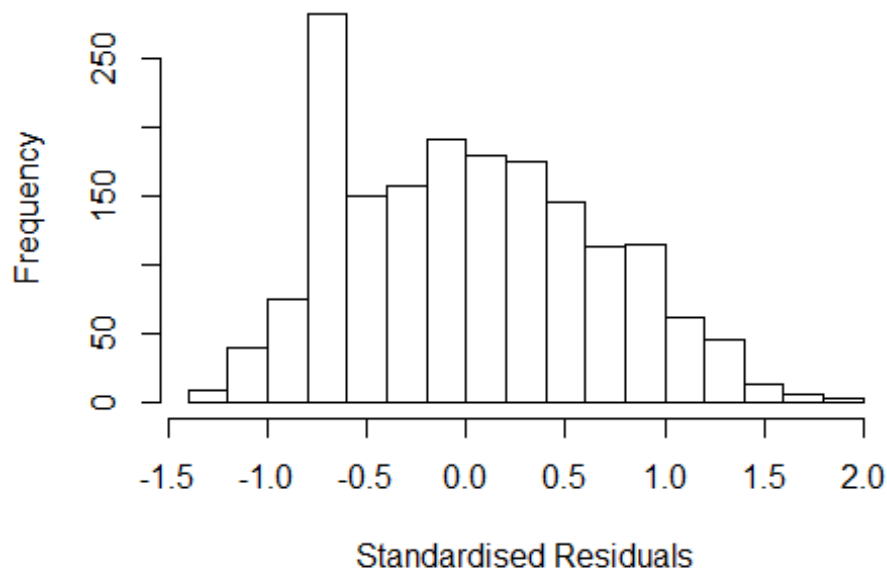
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4414 -0.5055 -0.0168  0.4518  1.8417
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3612     0.1001   13.60 < 2e-16 ***
## FirstAuthorFemale1  0.0660     0.0461    1.43  0.15245
## LastAuthorFemale1 -0.4342     0.0435   -9.99 < 2e-16 ***
## Year1997          -0.0820     0.1427   -0.57  0.56581
## Year1998          -0.0827     0.1384   -0.60  0.55034
## Year1999          -0.0430     0.1485   -0.29  0.77208
## Year2000          -0.0577     0.1273   -0.45  0.65046
## Year2001          -0.1518     0.1346   -1.13  0.25946
## Year2002           0.0143     0.1216    0.12  0.90667
## Year2003          -0.1251     0.1152   -1.09  0.27739
## Year2004          -0.1938     0.1220   -1.59  0.11234
## Year2005          -0.2462     0.1231   -2.00  0.04566 *
```

```

## Year2006          -0.4777      0.1163    -4.11  4.2e-05 ***
## Year2007          -0.4709      0.1171    -4.02  6.0e-05 ***
## Year2008          -0.4657      0.1173    -3.97  7.4e-05 ***
## Year2009          -0.4875      0.1102    -4.42  1.0e-05 ***
## Year2010          -0.4908      0.1131    -4.34  1.5e-05 ***
## Year2011          -0.5322      0.1122    -4.74  2.3e-06 ***
## Year2012          -0.4475      0.1174    -3.81  0.00014 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.629
## Multiple R-squared:  0.175, Adjusted R-squared:  0.167
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 136 weights are ~= 1. The remaining 1620 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.372  0.867   0.942   0.911   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.1 1      1.049
## Year              1.1 16      1.003

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3815 -0.5914 -0.0224 0.4828 1.8556
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3815 0.1096 12.60 < 2e-16 ***
## FirstAuthorFemale1 -0.1911 0.0359 -5.33 1.1e-07 ***
## Year1997 -0.1959 0.1549 -1.26 0.206
## Year1998 -0.0898 0.1525 -0.59 0.556
## Year1999 -0.0275 0.1671 -0.16 0.869
## Year2000 -0.0430 0.1398 -0.31 0.758
## Year2001 -0.1804 0.1487 -1.21 0.225
## Year2002 -0.0784 0.1325 -0.59 0.554
## Year2003 -0.1769 0.1258 -1.41 0.160
## Year2004 -0.2489 0.1324 -1.88 0.060 .
## Year2005 -0.3360 0.1312 -2.56 0.011 *
## Year2006 -0.5833 0.1244 -4.69 3.0e-06 ***
```

```

## Year2007          -0.5660      0.1266   -4.47  8.3e-06 ***
## Year2008          -0.5671      0.1255   -4.52  6.7e-06 ***
## Year2009          -0.5989      0.1198   -5.00  6.4e-07 ***
## Year2010          -0.5864      0.1232   -4.76  2.1e-06 ***
## Year2011          -0.6213      0.1210   -5.14  3.1e-07 ***
## Year2012          -0.5850      0.1266   -4.62  4.1e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.117, Adjusted R-squared:  0.108
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 112 weights are ~= 1. The remaining 1644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.381  0.864  0.927  0.909  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.69e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.191 1      1.091
## Year      1.191 16      1.005

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.393 -0.502 -0.012  0.461  1.780

```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3756    0.1000   13.76 < 2e-16 ***
## LastAuthorFemale1 -0.3964    0.0342  -11.59 < 2e-16 ***
## Year1997         -0.0812    0.1424   -0.57  0.56851
## Year1998         -0.0703    0.1384   -0.51  0.61127
## Year1999         -0.0310    0.1487   -0.21  0.83511
## Year2000         -0.0484    0.1271   -0.38  0.70327
## Year2001         -0.1407    0.1354   -1.04  0.29920
## Year2002          0.0170    0.1223    0.14  0.88947
## Year2003         -0.1122    0.1157   -0.97  0.33217
## Year2004         -0.1855    0.1221   -1.52  0.12882
## Year2005         -0.2445    0.1235   -1.98  0.04791 *
## Year2006         -0.4682    0.1166   -4.02  6.2e-05 ***
## Year2007         -0.4629    0.1180   -3.92  9.0e-05 ***
## Year2008         -0.4583    0.1178   -3.89  0.00010 ***
## Year2009         -0.4770    0.1110   -4.30  1.8e-05 ***
## Year2010         -0.4815    0.1140   -4.23  2.5e-05 ***
## Year2011         -0.5232    0.1129   -4.63  3.9e-06 ***
## Year2012         -0.4394    0.1182   -3.72  0.00021 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.634
## Multiple R-squared:  0.173, Adjusted R-squared:  0.165
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 133 weights are ~= 1. The remaining 1623 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.410  0.869  0.942  0.912  0.983  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.69e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1756"
## [1] ""

```

```

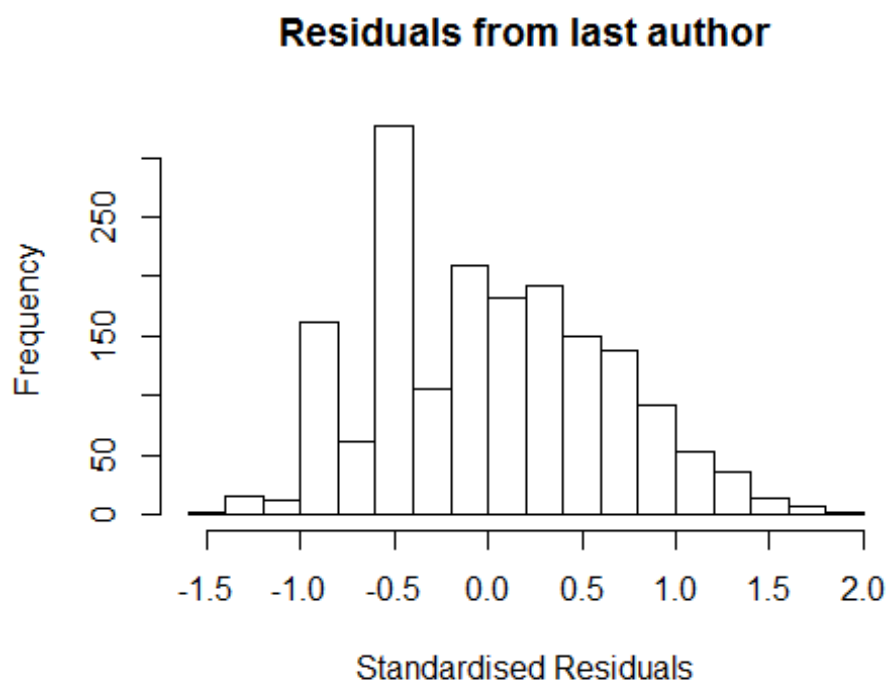
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2903"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    9    1    2    2    3    6    3    4    7    7    4    6    5    5    2
## 2012
##    2
##
## 1996 1997 1998 1999 2000 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    7    1    1    2    2    4    3    4    7    7    2    5    5    4    2
## 2012
##    2
##
## 1996 1997 1998 1999 2000 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    7    1    1    2    2    4    2    4    7    7    2    5    5    4    2
## 2012
##    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.71441761659491"
## [1] "Male first author team size 2018 geometric mean: 4"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.51984209978975"
## [1] "Male last author team size 2018 geometric mean: 5"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

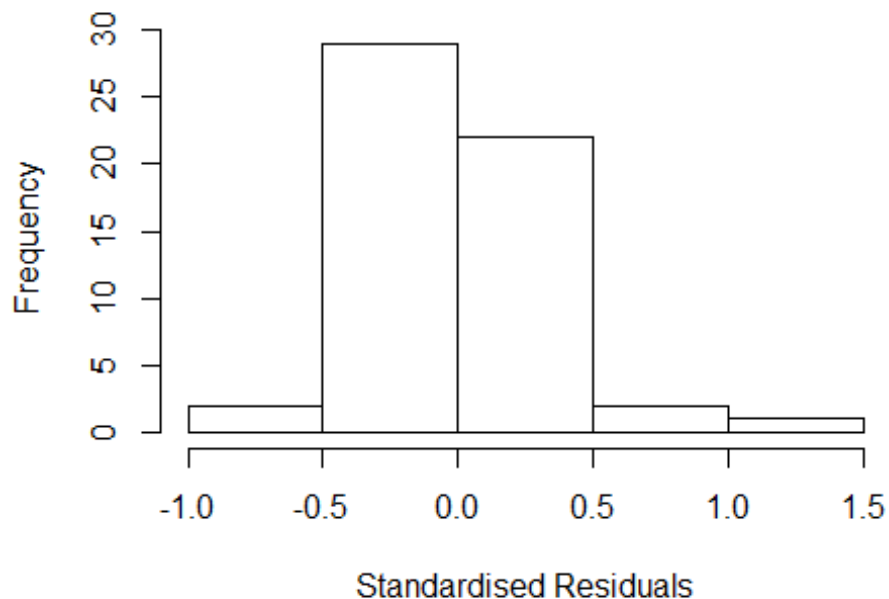
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
## FirstAuthorFemale	-9.026e+13	1	NaN
## LastAuthorFemale	1.814e+01	1	4.259
## UniqueAuthors	-4.977e+16	4	NaN
## Year	-7.839e+17	15	NaN

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9733 -0.1626 -0.0204 0.2330 1.4734
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5622 0.1969 2.85 0.0073 **
## FirstAuthorFemale1 -0.0967 0.1903 -0.51 0.6147
## LastAuthorFemale1 -0.2269 0.2824 -0.80 0.4273
## UniqueAuthors2 -0.0757 0.2374 -0.32 0.7518
## UniqueAuthors3 0.3775 0.2082 1.81 0.0786 .
## UniqueAuthors4 0.6245 0.3524 1.77 0.0853 .
## UniqueAuthors5 0.6272 0.2621 2.39 0.0224 *
## Year1997 -0.1932 0.1969 -0.98 0.3336
## Year1998 0.5068 0.1969 2.57 0.0146 *
## Year1999 -0.0315 0.3164 -0.10 0.9214
```

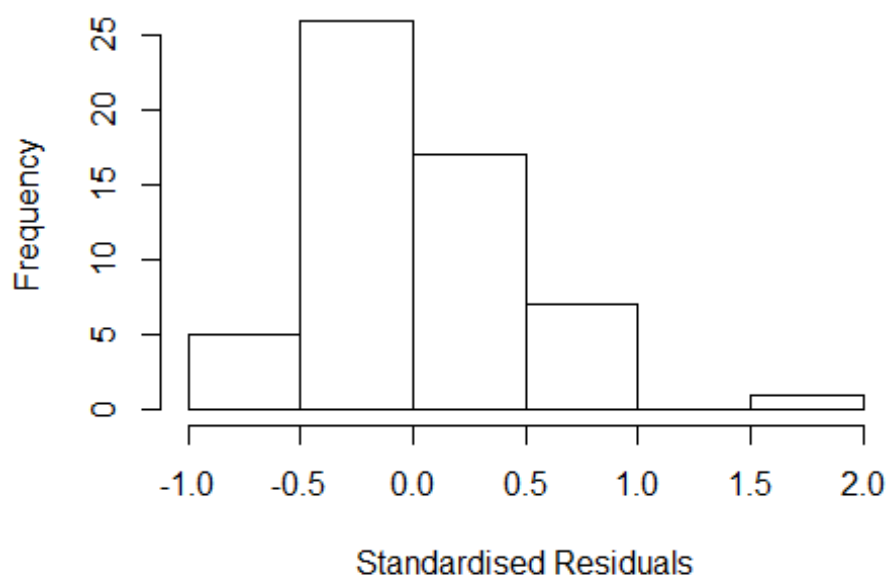


```

## Year2000          0.7576      0.3643      2.08      0.0452 *
## Year2002          0.6742      0.3262      2.07      0.0464 *
## Year2003         -0.0898      0.2850     -0.32      0.7546
## Year2004          0.4458      0.2732      1.63      0.1120
## Year2005          0.2342      0.2380      0.98      0.3322
## Year2006         -0.4871      0.2412     -2.02      0.0514 .
## Year2007          0.4940      0.4538      1.09      0.2841
## Year2008          0.4242      0.4608      0.92      0.3638
## Year2009         -0.0760      0.2746     -0.28      0.7836
## Year2010         -0.1099      0.2879     -0.38      0.7050
## Year2011         -0.3563      0.2483     -1.44      0.1604
## Year2012          0.6970      0.3299      2.11      0.0421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.348
## Multiple R-squared:  0.786, Adjusted R-squared:  0.653
## Convergence in 30 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 50 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0332 0.9040 0.9610 0.9070 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.79e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -1.393e+16 1      NaN
## LastAuthorFemale  1.796e+01 1      4.238
## Year              -5.114e+16 15      NaN

```

## Residuals from first and last author



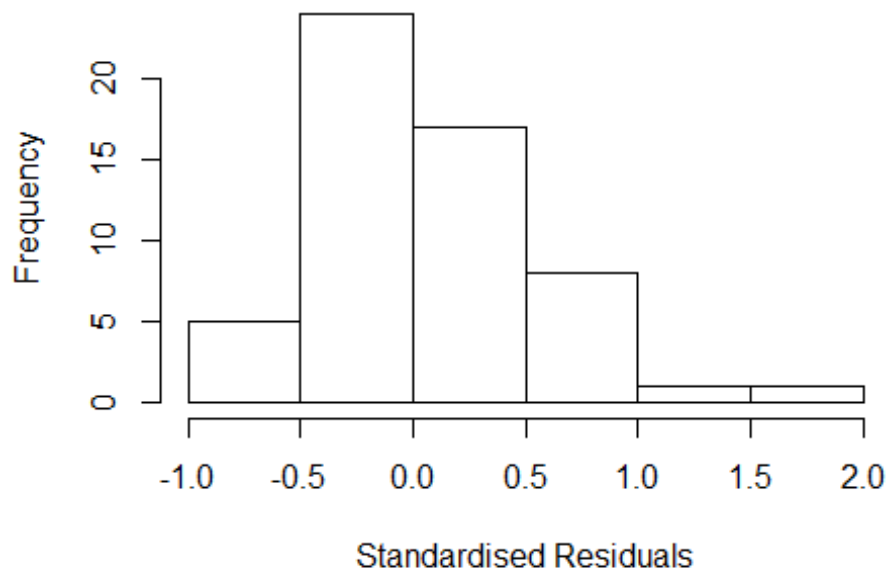
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.7431 -0.2198 -0.0239 0.2263 1.9532
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9107 0.1987 4.58 4.9e-05 ***
## FirstAuthorFemale1 -0.1075 0.3025 -0.36 0.7242
## LastAuthorFemale1 -0.7049 0.3053 -2.31 0.0265 *
## Year1997 -0.5417 0.1987 -2.73 0.0097 **
## Year1998 0.1583 0.1987 0.80 0.4307
## Year1999 0.1443 0.3486 0.41 0.6813
## Year2000 0.6848 0.2018 3.39 0.0016 **
## Year2002 0.9342 0.2874 3.25 0.0024 **
## Year2003 0.1238 0.3403 0.36 0.7180
## Year2004 0.6642 0.2275 2.92 0.0059 **
## Year2005 0.3381 0.2545 1.33 0.1920
## Year2006 -0.6909 0.2525 -2.74 0.0094 **
```

```

## Year2007          0.4266      0.9995      0.43      0.6720
## Year2008          0.4809      0.4259      1.13      0.2659
## Year2009          0.0965      0.3055      0.32      0.7539
## Year2010          0.1019      0.2631      0.39      0.7007
## Year2011         -0.2662      0.3220     -0.83      0.4136
## Year2012          0.9838      0.3440      2.86      0.0068 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.459
## Multiple R-squared:  0.669, Adjusted R-squared:  0.521
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 51 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.031  0.889   0.979   0.920   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.79e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.009e+15  1      3.176e+07
## Year              1.009e+15 15      3.163e+00

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -7.05e-01 -2.65e-01 1.11e-16 2.53e-01 1.91e+00
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.82813 0.13994 5.92 6.7e-07 ***
## FirstAuthorFemale1 -0.41837 0.29060 -1.44 0.15794
## Year1997 -0.45913 0.13994 -3.28 0.00219 **
## Year1998 0.24087 0.13994 1.72 0.09314 .
## Year1999 0.18525 0.65652 0.28 0.77931
## Year2000 0.76737 0.14427 5.32 4.6e-06 ***
## Year2002 1.08813 0.28257 3.85 0.00043 ***
## Year2003 0.16475 0.65315 0.25 0.80218
## Year2004 0.82722 0.18906 4.38 8.8e-05 ***
## Year2005 0.42188 0.21338 1.98 0.05513 .
## Year2006 -0.56265 0.30712 -1.83 0.07459 .
## Year2007 0.66456 0.56172 1.18 0.24394
```

```

## Year2008          0.81714      0.34432      2.37  0.02266 *
## Year2009          -0.01903      0.29916     -0.06  0.94960
## Year2010          -0.00841      0.23338     -0.04  0.97144
## Year2011          -0.22525      0.34666     -0.65  0.51965
## Year2012           1.37725      0.31975      4.31  0.00011 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.599, Adjusted R-squared:  0.435
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 47 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0976 0.8880 0.9650 0.9090 0.9840 0.9980
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.79e-03          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 5.187 1          2.277
## Year              5.187 15          1.056

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -0.7615 -0.2002 -0.0164  0.2136  1.9728
##

```

```

## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9106    0.2103   4.33  0.00010 ***
## LastAuthorFemale1 -0.7636    0.2191  -3.49  0.00123 **
## Year1997          -0.5416    0.2103  -2.58  0.01393 *
## Year1998           0.1584    0.2103   0.75  0.45574
## Year1999           0.0663    0.3184   0.21  0.83626
## Year2000           0.6849    0.2132   3.21  0.00264 **
## Year2002           0.9075    0.2817   3.22  0.00258 **
## Year2003           0.0458    0.3107   0.15  0.88368
## Year2004           0.6363    0.2314   2.75  0.00900 **
## Year2005           0.3374    0.2636   1.28  0.20811
## Year2006          -0.7103    0.2522  -2.82  0.00759 **
## Year2007           0.3729    1.4583   0.26  0.79950
## Year2008           0.3872    0.3395   1.14  0.26100
## Year2009           0.0636    0.3330   0.19  0.84954
## Year2010           0.0781    0.2888   0.27  0.78811
## Year2011          -0.3442    0.3123  -1.10  0.27716
## Year2012           0.8764    0.2103   4.17  0.00017 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.442
## Multiple R-squared:  0.675, Adjusted R-squared:  0.541
## Convergence in 29 IRWLS iterations
##
## Robustness weights:
## 4 weights are ~= 1. The remaining 52 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0084 0.8850 0.9790 0.9160 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.79e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500            50          2            1            1000      200
##      trace.lev      mts      compute.rd
##      0              1000      0
##      psi            subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 56"
## [1] ""
## [1] ""
## [1] "#####"

```

```

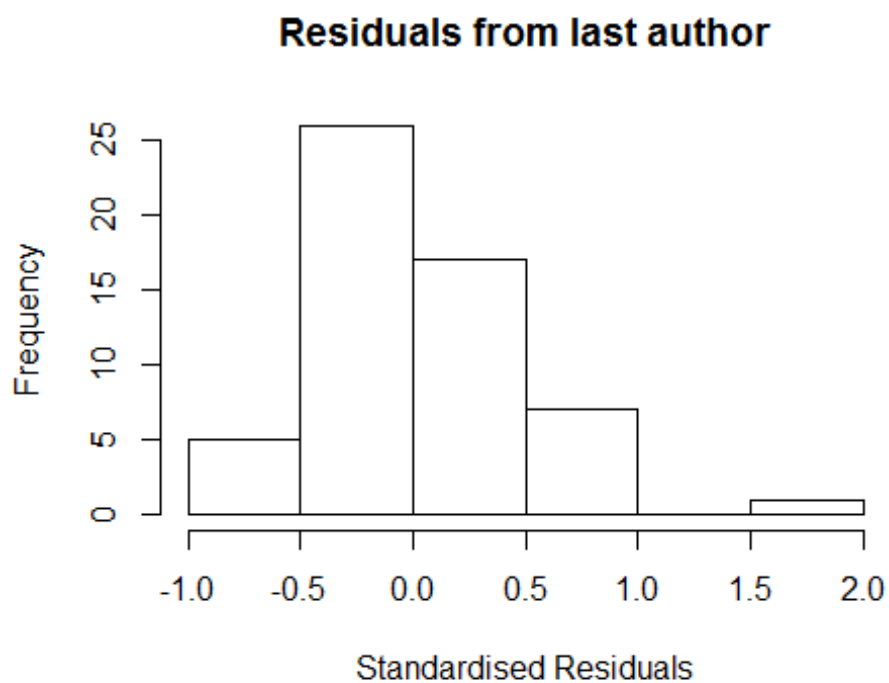
## [1] "Analysis of AJSC 2904"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    1    2   12   15   24   20   17   14   22   32   26   30   19    9   18
##
## 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    1   11   15   24   18   15   14   18   28   22   28   16    9   17
##
## 1997 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
##    0    1   10   15   22   17   14   13   18   25   21   27   15    8   15
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.06844987990576"
## [1] "Male first author team size 2018 geometric mean: 1.49534878122122"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.71877192758748"
## [1] "Male last author team size 2018 geometric mean: 1.94416129723967"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

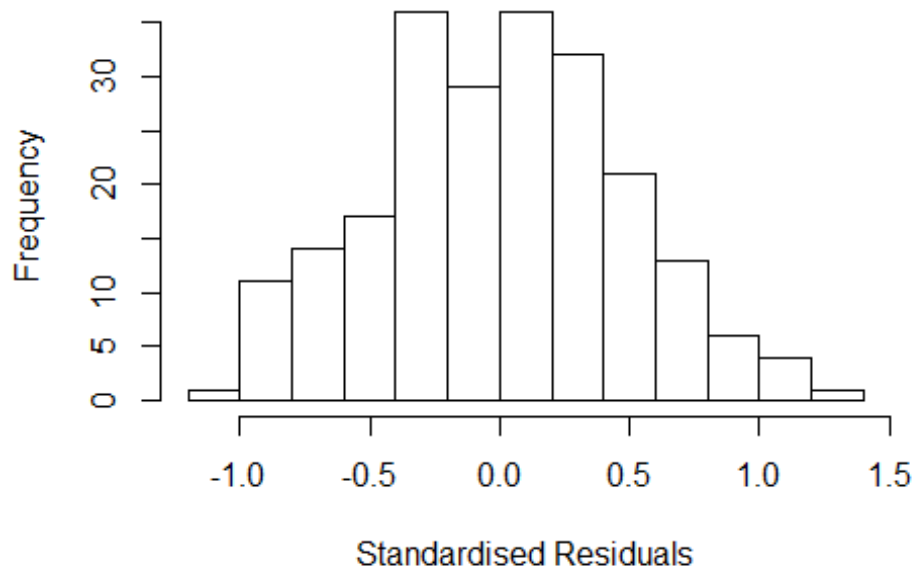
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.176 1      1.475
## LastAuthorFemale  2.938 1      1.714
## UniqueAuthors    12.225 4      1.367
## Year              26.577 13     1.134
```



## Residuals from first and last author and team size



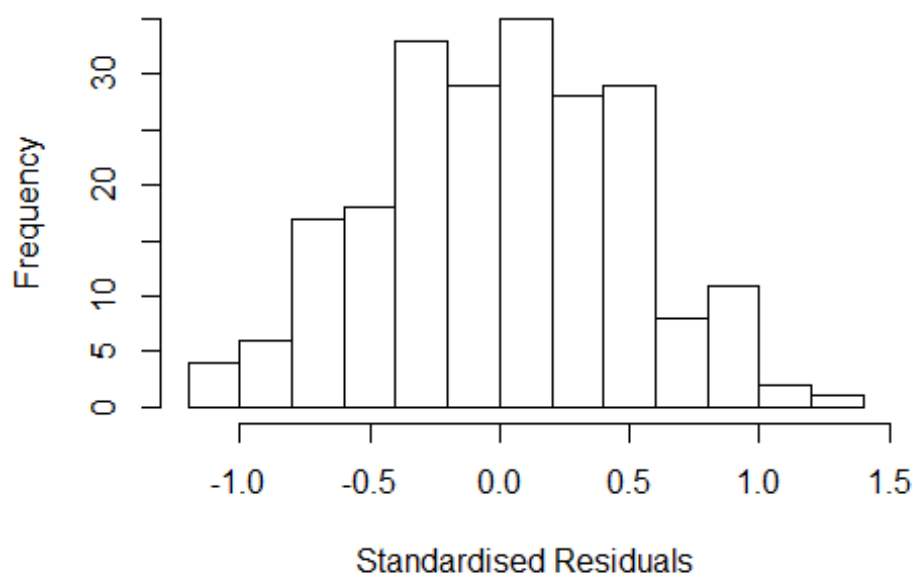
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.05260 -0.33476 0.00316 0.33017 1.29224
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0547 0.1477 -0.37 0.71157
## FirstAuthorFemale1 0.0193 0.0849 0.23 0.82025
## LastAuthorFemale1 0.0650 0.0793 0.82 0.41363
## UniqueAuthors2 -0.0296 0.0977 -0.30 0.76216
## UniqueAuthors3 0.0315 0.1134 0.28 0.78112
## UniqueAuthors4 0.1570 0.1167 1.35 0.18007
## UniqueAuthors5 0.1704 0.1244 1.37 0.17210
## Year2000 0.7483 0.2526 2.96 0.00342 **
## Year2001 0.7309 0.1568 4.66 5.7e-06 ***
## Year2002 0.9915 0.1284 7.72 5.4e-13 ***
```

```

## Year2003          0.7332      0.1478      4.96 1.5e-06 ***
## Year2004          0.9894      0.0974     10.15 < 2e-16 ***
## Year2005          0.8535      0.1601      5.33 2.6e-07 ***
## Year2006          0.8045      0.1817      4.43 1.6e-05 ***
## Year2007          0.5992      0.1265      4.74 4.1e-06 ***
## Year2008          0.9127      0.1625      5.62 6.4e-08 ***
## Year2009          0.5975      0.1362      4.39 1.8e-05 ***
## Year2010          0.6050      0.1557      3.88 0.00014 ***
## Year2011          0.4134      0.2060      2.01 0.04615 *
## Year2012          0.4680      0.1362      3.43 0.00072 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.144, Adjusted R-squared:  0.0628
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 24 weights are ~ = 1. The remaining 197 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.512  0.881  0.957  0.917  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.768 1      1.330
## LastAuthorFemale  2.801 1      1.674
## Year              3.596 13      1.050

```

## Residuals from first and last author



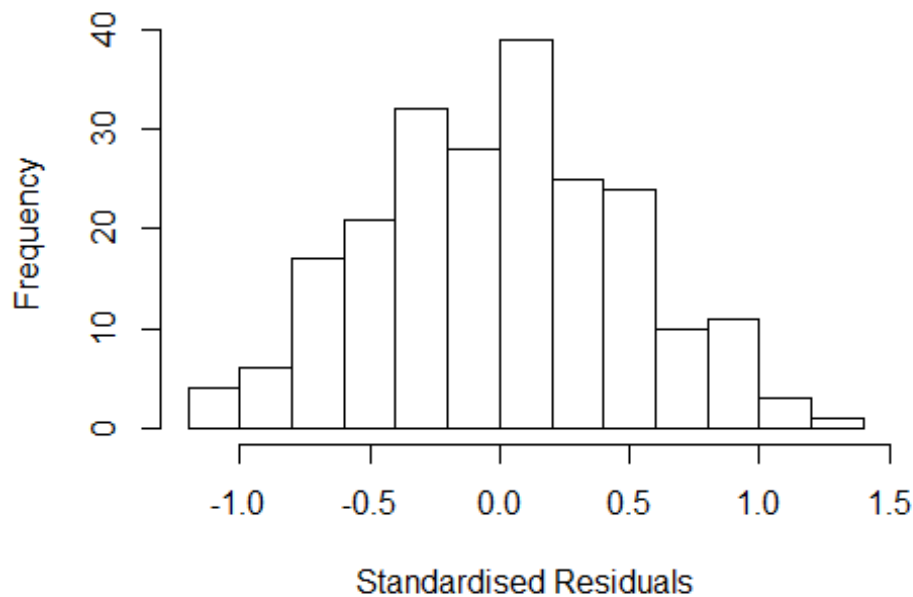
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0927 -0.3446 0.0123 0.3454 1.2265
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.1063 0.1000 -1.06 0.2888
## FirstAuthorFemale1 0.0371 0.0808 0.46 0.6465
## LastAuthorFemale1 0.0692 0.0808 0.86 0.3924
## Year2000 0.7926 0.2754 2.88 0.0044 **
## Year2001 0.8086 0.1318 6.14 4.3e-09 ***
## Year2002 1.0927 0.1091 10.01 < 2e-16 ***
## Year2003 0.8285 0.1517 5.46 1.4e-07 ***
## Year2004 1.0463 0.0814 12.86 < 2e-16 ***
## Year2005 0.8954 0.1339 6.68 2.1e-10 ***
## Year2006 0.8864 0.1724 5.14 6.3e-07 ***
## Year2007 0.6846 0.1013 6.76 1.4e-10 ***
## Year2008 1.0256 0.1425 7.20 1.1e-11 ***
```

```

## Year2009          0.6916      0.1092      6.33  1.5e-09 ***
## Year2010          0.7359      0.1325      5.55  8.6e-08 ***
## Year2011          0.4861      0.1778      2.73   0.0068 **
## Year2012          0.5223      0.1225      4.27  3.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.508
## Multiple R-squared:  0.126, Adjusted R-squared:  0.0625
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 199 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.540  0.882  0.950  0.913  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.651 1          1.285
## Year              1.651 13          1.019

```

## Residuals from first author



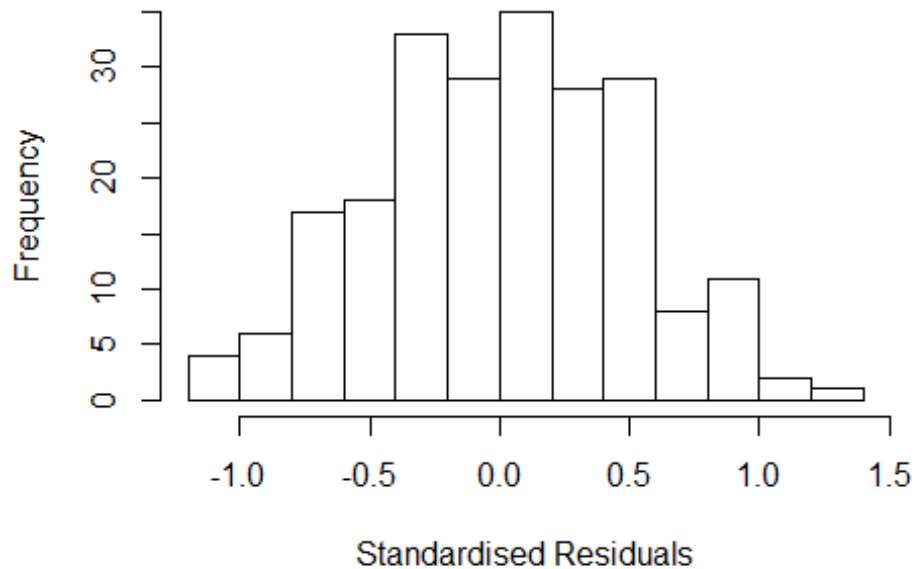
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0553 -0.3429  0.0284  0.3301  1.2564
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.0536    0.0786   -0.68   0.4959
## FirstAuthorFemale1  0.0536    0.0786    0.68   0.4959
## Year2000        0.7582    0.2562    2.96   0.0034 **
## Year2001        0.7745    0.1216    6.37  1.2e-09 ***
## Year2002        1.0553    0.0992   10.64 < 2e-16 ***
## Year2003        0.7986    0.1504    5.31  2.8e-07 ***
## Year2004        1.0382    0.0844   12.30 < 2e-16 ***
## Year2005        0.8690    0.1308    6.64  2.7e-10 ***
## Year2006        0.8566    0.1701    5.04  1.0e-06 ***
## Year2007        0.6806    0.1002    6.79  1.2e-10 ***
## Year2008        1.0003    0.1346    7.43  2.8e-12 ***
## Year2009        0.6454    0.0900    7.17  1.3e-11 ***
```

```

## Year2010          0.7159      0.1310      5.47  1.3e-07 ***
## Year2011          0.4529      0.1768      2.56   0.0111 *
## Year2012          0.5088      0.1252      4.06  6.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.122, Adjusted R-squared:  0.0628
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.524  0.874  0.946  0.914  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.52e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 2.581 1          1.606
## Year            2.581 13          1.037

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0813 -0.3456 0.0237 0.3557 1.2309
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0782 0.0784 -1.00 0.3197
## LastAuthorFemale1 0.0782 0.0784 1.00 0.3197
## Year2000 0.7838 0.2706 2.90 0.0042 **
## Year2001 0.8016 0.1310 6.12 4.6e-09 ***
## Year2002 1.0813 0.1062 10.18 < 2e-16 ***
## Year2003 0.8241 0.1520 5.42 1.7e-07 ***
## Year2004 1.0448 0.0812 12.87 < 2e-16 ***
## Year2005 0.8892 0.1333 6.67 2.3e-10 ***
## Year2006 0.8774 0.1705 5.15 6.2e-07 ***
## Year2007 0.6743 0.0998 6.75 1.4e-10 ***
## Year2008 1.0180 0.1420 7.17 1.3e-11 ***
## Year2009 0.6857 0.1106 6.20 3.1e-09 ***
```

```

## Year2010          0.7210      0.1263      5.71  3.9e-08 ***
## Year2011          0.4684      0.1748      2.68  0.0079 **
## Year2012          0.5084      0.1193      4.26  3.1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.51
## Multiple R-squared:  0.125, Adjusted R-squared:  0.0657
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 200 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.540  0.875  0.949  0.914  0.983  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          4.52e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 221"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2905"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1   33    1   83  107  100   35   30   34   51   72   82  135  159  173
## 2012
##   165
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1   32    1   45   50   85   27   25   32   45   64   76  127  143  157
## 2012
##   150
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011

```



```

##      1      31      1      43      50      80      25      23      30      45      56      66      113      126      142
## 2012
## 138
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.5252963915611"
## [1] "Male first author team size 2018 geometric mean: 1.35299843866192"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2800, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.46336920921948"
## [1] "Male last author team size 2018 geometric mean: 1.50350880000367"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2600, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.353e+00 1          1.831
## LastAuthorFemale  3.250e+00 1          1.803
## UniqueAuthors    1.088e+11 4          23.964
## Year              1.975e+11 15         2.380
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2510 -0.4150 -0.0316  0.4765  2.3931
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1884      0.0466   25.48 < 2e-16 ***
## FirstAuthorFemale1 0.1371      0.0596    2.30  0.0216 *
```

```

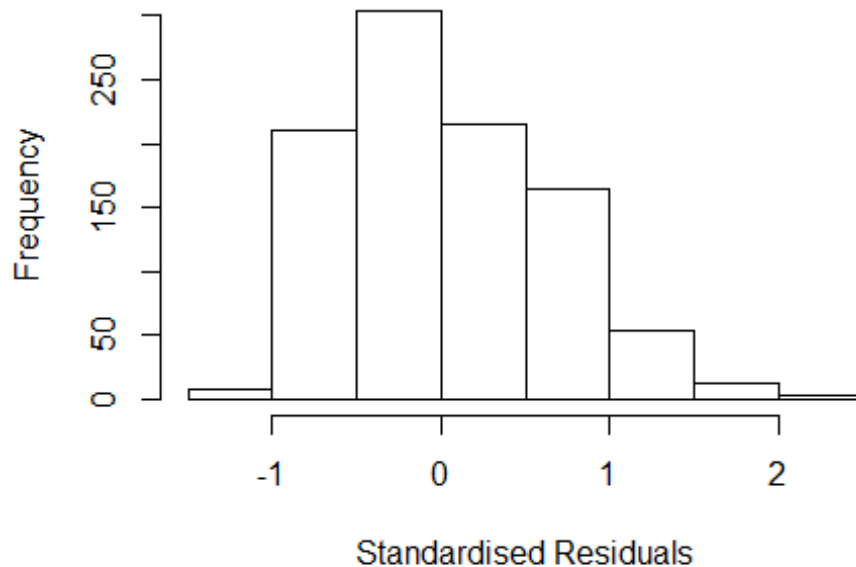
## LastAuthorFemale1 -0.0605 0.0585 -1.04 0.3007
## UniqueAuthors2 0.2789 0.0545 5.12 3.7e-07 ***
## UniqueAuthors3 0.3247 0.1098 2.96 0.0032 **
## UniqueAuthors4 0.4147 0.0922 4.50 7.8e-06 ***
## UniqueAuthors5 0.5817 0.1422 4.09 4.7e-05 ***
## Year1998 -0.9312 0.0967 -9.63 < 2e-16 ***
## Year1999 0.0471 0.0545 0.86 0.3874
## Year2000 -0.8770 0.0810 -10.82 < 2e-16 ***
## Year2001 -0.7338 0.0794 -9.24 < 2e-16 ***
## Year2002 -0.5661 0.0967 -5.86 6.5e-09 ***
## Year2003 -0.5169 0.1264 -4.09 4.7e-05 ***
## Year2004 -0.4605 0.1533 -3.00 0.0027 **
## Year2005 -0.6705 0.1152 -5.82 8.0e-09 ***
## Year2006 -0.6410 0.0972 -6.59 7.1e-11 ***
## Year2007 -0.8784 0.0892 -9.85 < 2e-16 ***
## Year2008 -0.8621 0.0827 -10.42 < 2e-16 ***
## Year2009 -0.8500 0.0614 -13.85 < 2e-16 ***
## Year2010 -0.6468 0.0645 -10.03 < 2e-16 ***
## Year2011 -0.5191 0.0732 -7.10 2.5e-12 ***
## Year2012 -0.7254 0.0602 -12.05 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared: 0.12, Adjusted R-squared: 0.101
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 67 weights are ~= 1. The remaining 903 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.0859 0.8800 0.9390 0.9070 0.9750 0.9990
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.03e-04 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
## 500 50 2 1 1000 200
## trace.lev mts compute.rd
## 0 1000 0
## psi subsampling cov
## "bisquare" "nonsingular" ".vcov.avar1"
## compute.outlier.stats
## "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

```

```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```

## Residuals from first and last author and team size



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1          NaN
## LastAuthorFemale  NaN 1          NaN
## Year              NaN 15         NaN

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9428 -0.5278 -0.0241  0.4967  2.3118
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1734     0.0474   24.73 < 2e-16 ***
```

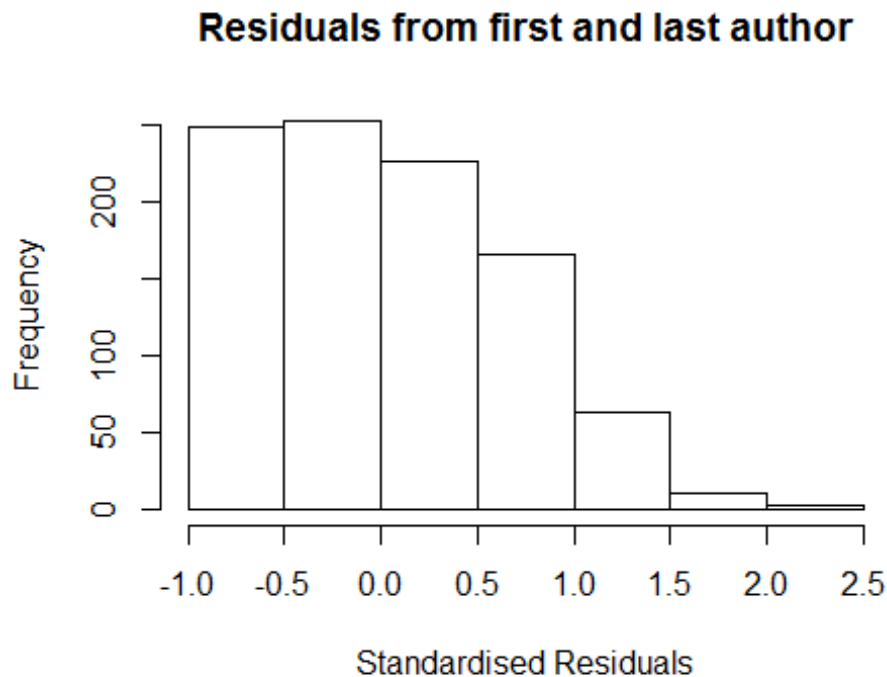
```

## FirstAuthorFemale1    0.1360    0.0622    2.19    0.029 *
## LastAuthorFemale1    -0.0443    0.0616   -0.72    0.472
## Year1998              -0.8803    0.0986   -8.93 < 2e-16 ***
## Year1999              0.3260    0.0000    Inf < 2e-16 ***
## Year2000              -0.8225    0.0781  -10.52 < 2e-16 ***
## Year2001              -0.6422    0.0769   -8.35 2.3e-16 ***
## Year2002              -0.4056    0.0932   -4.35 1.5e-05 ***
## Year2003              -0.3455    0.1391   -2.48 0.013 *
## Year2004              -0.3222    0.1633   -1.97 0.049 *
## Year2005              -0.5552    0.1234   -4.50 7.6e-06 ***
## Year2006              -0.5338    0.0939   -5.69 1.7e-08 ***
## Year2007              -0.7797    0.0912   -8.55 < 2e-16 ***
## Year2008              -0.7354    0.0922   -7.98 4.3e-15 ***
## Year2009              -0.7372    0.0614  -12.00 < 2e-16 ***
## Year2010              -0.5488    0.0649   -8.46 < 2e-16 ***
## Year2011              -0.3809    0.0684   -5.57 3.3e-08 ***
## Year2012              -0.5893    0.0565  -10.44 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.654
## Multiple R-squared:  0.0619, Adjusted R-squared:  0.0451
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 908 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.185  0.891  0.941  0.915  0.978  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.03e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

```

```
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```



```
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1             NaN
## Year            NaN 15             NaN

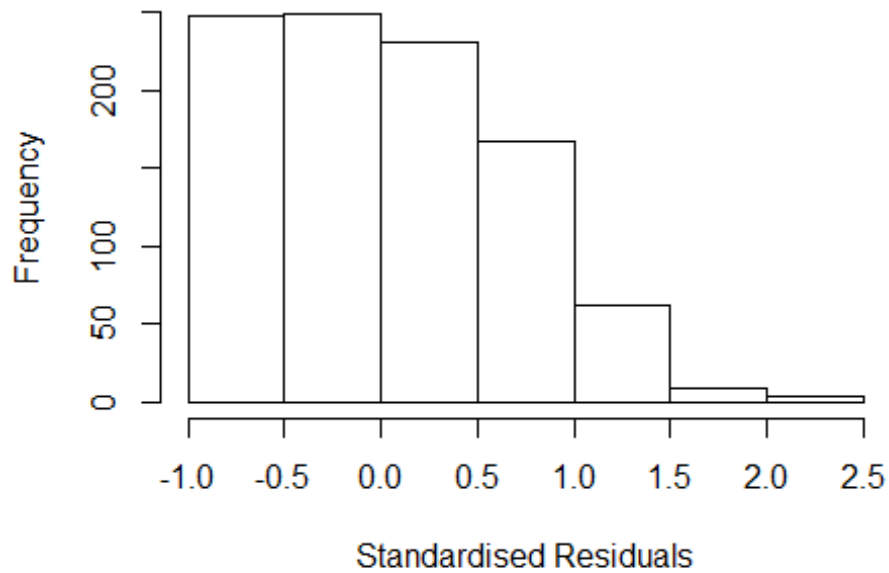
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9424 -0.5317 -0.0255  0.4816  2.3087
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1602     0.0439   26.43 < 2e-16 ***
## FirstAuthorFemale1 0.1048     0.0439    2.39  0.017 *
## Year1998         -0.8735     0.0976   -8.95 < 2e-16 ***
## Year1999          0.3260     0.0000    Inf < 2e-16 ***
## Year2000         -0.8178     0.0785  -10.42 < 2e-16 ***
```

```

## Year2001      -0.6351      0.0765      -8.30      3.5e-16 ***
## Year2002      -0.4010      0.0925      -4.34      1.6e-05 ***
## Year2003      -0.3404      0.1390      -2.45       0.015 *
## Year2004      -0.3226      0.1630      -1.98       0.048 *
## Year2005      -0.5525      0.1228      -4.50      7.6e-06 ***
## Year2006      -0.5292      0.0930      -5.69      1.7e-08 ***
## Year2007      -0.7753      0.0908      -8.54      < 2e-16 ***
## Year2008      -0.7333      0.0925      -7.93      6.2e-15 ***
## Year2009      -0.7329      0.0603     -12.16      < 2e-16 ***
## Year2010      -0.5457      0.0652      -8.37      < 2e-16 ***
## Year2011      -0.3785      0.0683      -5.54      3.9e-08 ***
## Year2012      -0.5869      0.0563     -10.42      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0614, Adjusted R-squared:  0.0457
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 909 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.182  0.891  0.940  0.914  0.979  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.03e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful

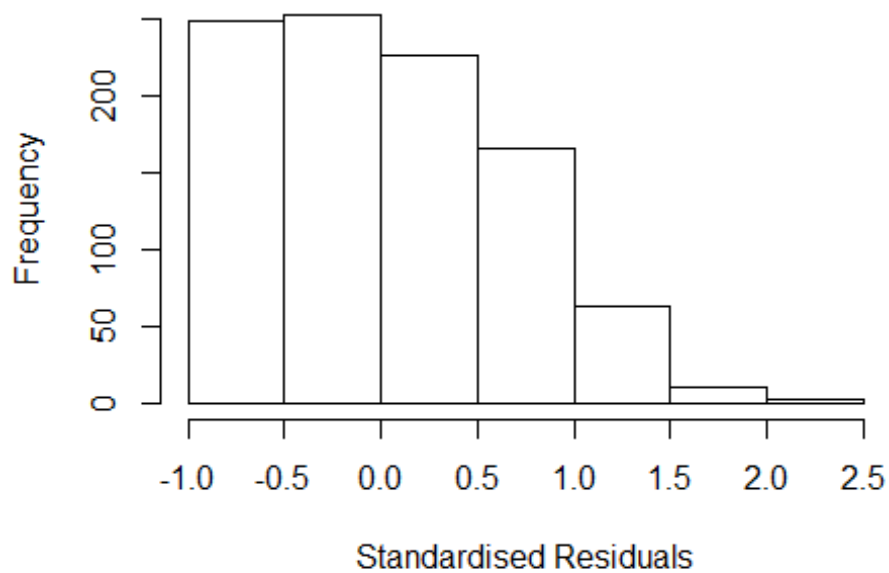
```

### Residuals from first author



##		GVIF	Df	$GVIF^{(1/(2*Df))}$
##	LastAuthorFemale	NaN	1	NaN
##	Year	NaN	15	NaN

### Residuals from last author



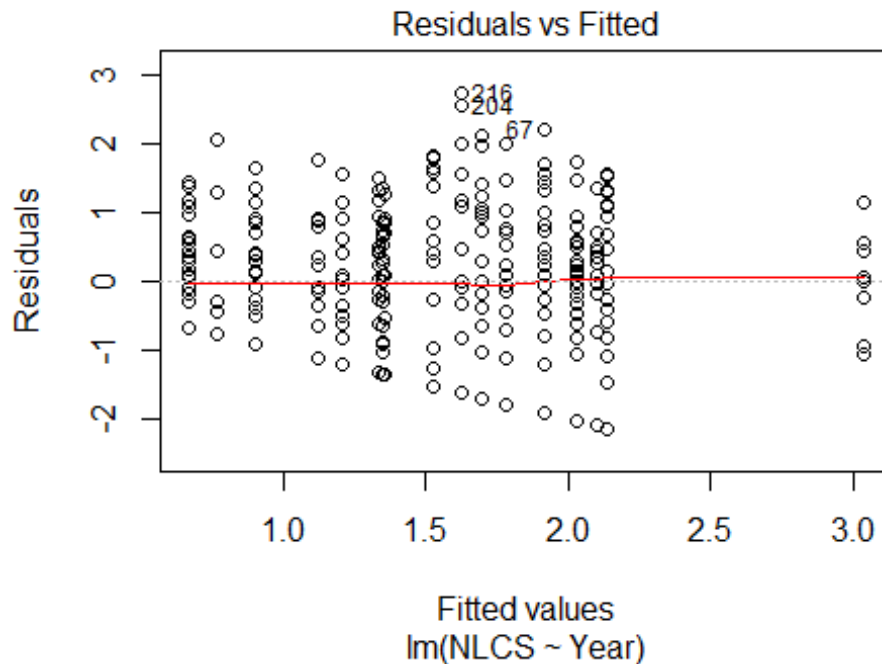
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9144 -0.5148 -0.0354 0.4807 2.3320
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2137 0.0441 27.51 < 2e-16 ***
## LastAuthorFemale1 0.0513 0.0441 1.16 0.245
## Year1998 -0.8940 0.0998 -8.96 < 2e-16 ***
## Year1999 0.3260 0.0000 Inf < 2e-16 ***
## Year2000 -0.8313 0.0784 -10.61 < 2e-16 ***
## Year2001 -0.6447 0.0763 -8.45 < 2e-16 ***
## Year2002 -0.4113 0.0927 -4.43 1.0e-05 ***
## Year2003 -0.3506 0.1385 -2.53 0.012 *
## Year2004 -0.3588 0.1610 -2.23 0.026 *
## Year2005 -0.5740 0.1198 -4.79 1.9e-06 ***
## Year2006 -0.5544 0.0935 -5.93 4.3e-09 ***
## Year2007 -0.7881 0.0908 -8.68 < 2e-16 ***
## Year2008 -0.7461 0.0931 -8.02 3.2e-15 ***
## Year2009 -0.7502 0.0611 -12.27 < 2e-16 ***
## Year2010 -0.5690 0.0631 -9.02 < 2e-16 ***
## Year2011 -0.4052 0.0671 -6.04 2.2e-09 ***
## Year2012 -0.6051 0.0565 -10.70 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.652
## Multiple R-squared: 0.057, Adjusted R-squared: 0.0411
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 64 weights are ~= 1. The remaining 906 ones are summarized as
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.173 0.887 0.944 0.914 0.979 0.999
## Algorithmic parameters:
## tuning.chi bb tuning.psi refine.tol
## 1.55e+00 5.00e-01 4.69e+00 1.00e-07
## rel.tol solve.tol eps.outlier eps.x
## 1.00e-07 1.00e-07 1.03e-04 1.82e-12
## warn.limit.reject warn.limit.meanrw
## 5.00e-01 5.00e-01
```



```

##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##          500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##          0          1000          0
##          psi          subsampling          cov
##          "bisquare"          "nonsingular"          ".vcov.avar1"
##      compute.outlier.stats
##          "SM"
##      seed : int(0)
## [1] "Sample size for the above analysis: 970"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2906"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   30   48   34   29   24   25   27   35   34   20   27   22   19   10   44
## 2011 2012
##   35   48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   35   25   22    8   11   22   22   29   18   24   19   15    9   42
## 2011 2012
##   16   33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   24   34   24   21    8   10   21   22   26   16   22   15   14    9   34
## 2011 2012
##   12   30
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
##      Bartlett test of homogeneity of variances
##
##      data:  NLCS by Year
##      Bartlett's K-squared = 34, df = 16, p-value = 0.005

```



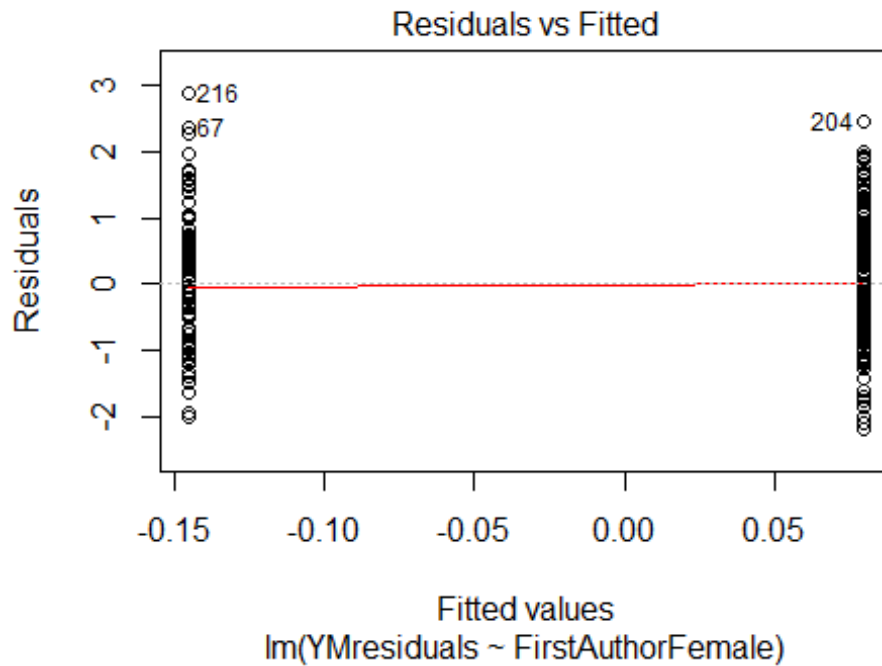
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.095, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 3.54349832953353"
## [1] "Male first author team size 2018 geometric mean: 3.18416396911442"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

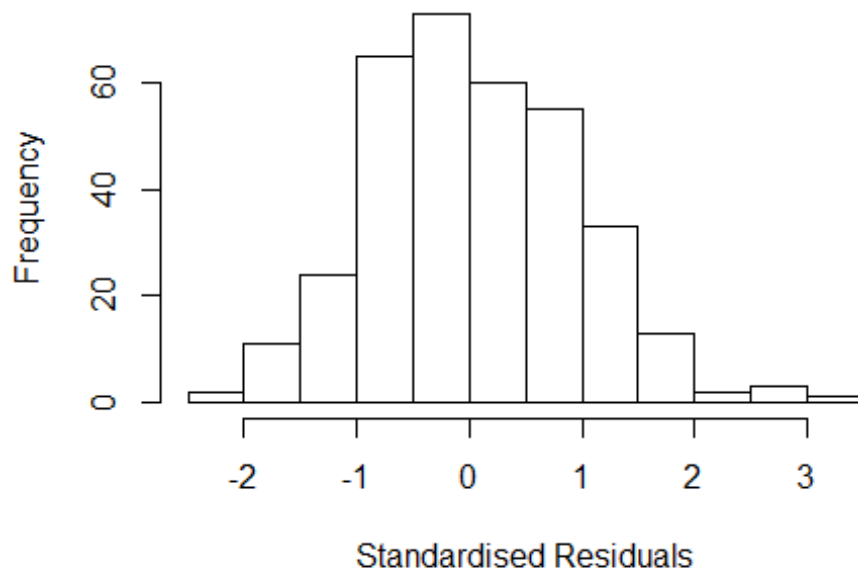
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.4519247197843"
## [1] "Male last author team size 2018 geometric mean: 3.27773137679653"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 96, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
##      Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.389 1          1.545
## LastAuthorFemale  2.260 1          1.503
## UniqueAuthors    2.768 4          1.136
## Year              4.157 16          1.046
```

## Residuals from first and last author and team size



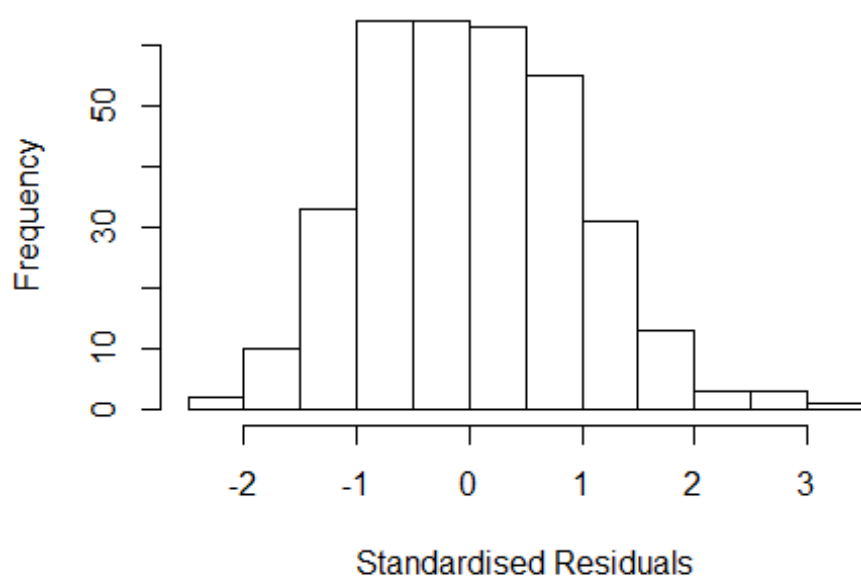
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 16  0030115207 3.826 1996    2906      1    2.576
## 67  0031242183 4.117 1997    2906      1    2.539
## 204 0036011010 4.168 2002    2906      1    2.878
## 216 0036702696 4.354 2002    2906      1    3.411
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1248 -0.6183 -0.0246  0.6600  3.4110
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.250417   0.254955   4.90  1.5e-06 ***
## FirstAuthorFemale1 0.135838   0.154888   0.88  0.38114
## LastAuthorFemale1 0.319394   0.146519   2.18  0.03000 *
## UniqueAuthors2    0.211124   0.153971   1.37  0.17128
## UniqueAuthors3   -0.000211   0.159826   0.00  0.99895
## UniqueAuthors4    0.000234   0.207061   0.00  0.99910
## UniqueAuthors5    0.305946   0.210247   1.46  0.14661
```

```

## Year1997          0.327736    0.294328    1.11  0.26633
## Year1998          0.594496    0.308874    1.92  0.05515 .
## Year1999          0.136955    0.329730    0.42  0.67816
## Year2000          1.362605    0.359205    3.79  0.00018 ***
## Year2001         -0.303107    0.484182   -0.63  0.53175
## Year2002         -0.307442    0.460306   -0.67  0.50468
## Year2003          0.527422    0.290023    1.82  0.06992 .
## Year2004          0.209259    0.267107    0.78  0.43396
## Year2005          0.142578    0.487876    0.29  0.77029
## Year2006         -0.271429    0.297973   -0.91  0.36303
## Year2007         -0.347281    0.359499   -0.97  0.33477
## Year2008         -0.499195    0.360709   -1.38  0.16735
## Year2009         -0.957796    0.394559   -2.43  0.01576 *
## Year2010         -0.750713    0.272722   -2.75  0.00625 **
## Year2011         -0.273668    0.361690   -0.76  0.44983
## Year2012         -0.632113    0.279465   -2.26  0.02438 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.889
## Multiple R-squared:  0.289, Adjusted R-squared:  0.24
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 315 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.872  0.946  0.904  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.92e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.347 1          1.532
## LastAuthorFemale 2.144 1          1.464
## Year              1.602 16          1.015

```

## Residuals from first and last author



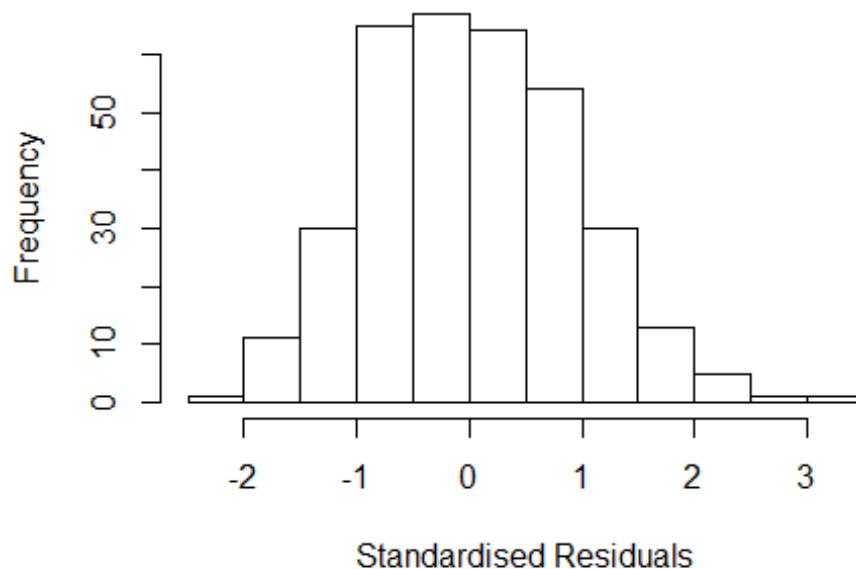
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 16  0030115207 3.826 1996    2906      1    2.554
## 67  0031242183 4.117 1997    2906      1    2.510
## 204 0036011010 4.168 2002    2906      1    2.962
## 216 0036702696 4.354 2002    2906      1    3.303
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1670 -0.6289 -0.0277  0.6309  3.3027
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.272      0.255   4.99 9.9e-07 ***
## FirstAuthorFemale1  0.155      0.154   1.01 0.31477
## LastAuthorFemale1  0.294      0.144   2.04 0.04232 *
## Year1997          0.335      0.294   1.14 0.25521
## Year1998          0.623      0.312   2.00 0.04660 *
## Year1999          0.200      0.326   0.61 0.53979
## Year2000          1.383      0.357   3.87 0.00013 ***
## Year2001         -0.275      0.500  -0.55 0.58298
## Year2002         -0.221      0.442  -0.50 0.61692
```

```

## Year2003          0.600      0.284      2.12  0.03506 *
## Year2004          0.248      0.264      0.94  0.34906
## Year2005          0.156      0.502      0.31  0.75667
## Year2006         -0.246      0.300     -0.82  0.41251
## Year2007         -0.197      0.340     -0.58  0.56279
## Year2008         -0.406      0.347     -1.17  0.24302
## Year2009         -0.969      0.416     -2.33  0.02036 *
## Year2010         -0.697      0.263     -2.65  0.00844 **
## Year2011         -0.198      0.346     -0.57  0.56662
## Year2012         -0.579      0.271     -2.13  0.03363 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.9
## Multiple R-squared:  0.279, Adjusted R-squared:  0.239
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.149  0.875   0.949   0.905   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.92e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.519 1          1.232
## Year              1.519 16          1.013

```

## Residuals from first author



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 16  0030115207 3.826 1996    2906      1    2.554
## 67  0031242183 4.117 1997    2906      1    2.510
## 204 0036011010 4.168 2002    2906      1    2.962
## 216 0036702696 4.354 2002    2906      1    3.303
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2665 -0.6490 -0.0109  0.6431  3.1879
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.3613    0.2580     5.28 2.4e-07 ***
## FirstAuthorFemale1 0.3327    0.1245     2.67 0.00792 **
## Year1997        0.3059    0.2982     1.03 0.30571
## Year1998        0.6117    0.3242     1.89 0.06003 .
## Year1999        0.1601    0.3267     0.49 0.62451
## Year2000        1.3943    0.3578     3.90 0.00012 ***
## Year2001       -0.2797    0.4841    -0.58 0.56386
## Year2002       -0.1952    0.4637    -0.42 0.67409
## Year2003        0.5725    0.2876     1.99 0.04736 *
```

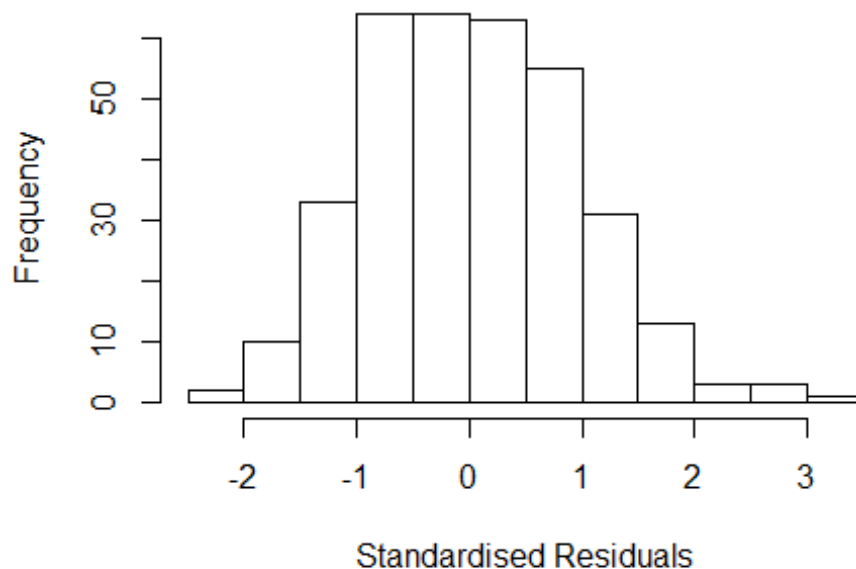


```

## Year2004          0.2661      0.2696      0.99  0.32436
## Year2005          0.0792      0.5038      0.16  0.87517
## Year2006         -0.3055      0.2960     -1.03  0.30291
## Year2007         -0.2874      0.3428     -0.84  0.40237
## Year2008         -0.4854      0.3688     -1.32  0.18910
## Year2009         -1.0147      0.4071     -2.49  0.01318 *
## Year2010         -0.7323      0.2668     -2.74  0.00639 **
## Year2011         -0.2863      0.3291     -0.87  0.38494
## Year2012         -0.6413      0.2773     -2.31  0.02136 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.9
## Multiple R-squared:  0.268, Adjusted R-squared:  0.229
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 319 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.184  0.869  0.950   0.906   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.92e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.314 1      1.146
## Year            1.314 16      1.009

```

## Residuals from last author



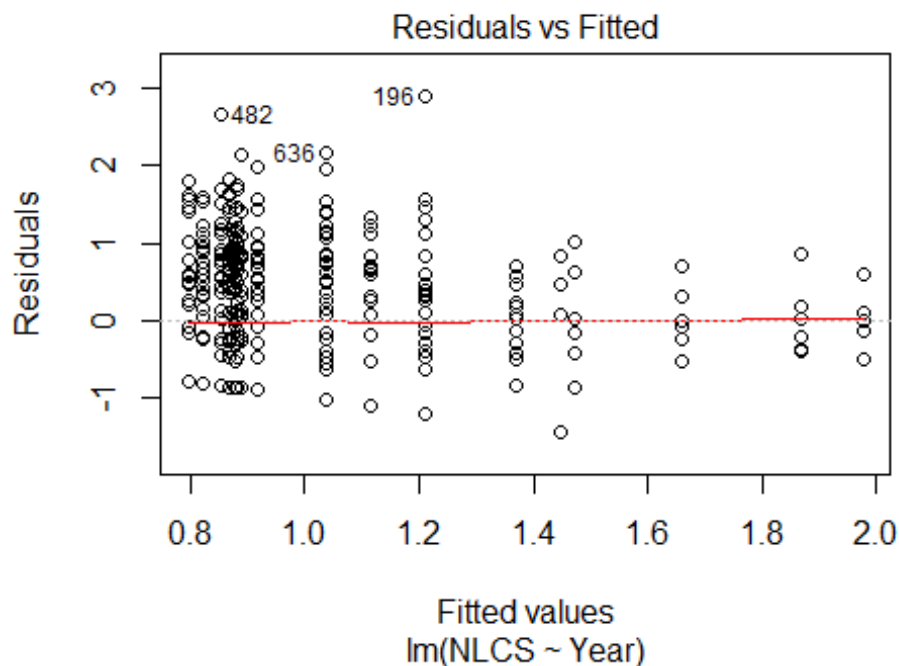
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 16  0030115207 3.826 1996    2906      1    2.554
## 67  0031242183 4.117 1997    2906      1    2.510
## 204 0036011010 4.168 2002    2906      1    2.962
## 216 0036702696 4.354 2002    2906      1    3.303
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.315 -0.608 -0.018  0.636  3.265
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.332      0.251   5.30 2.1e-07 ***
## LastAuthorFemale1  0.377      0.114   3.31 0.00103 **
## Year1997          0.320      0.291   1.10 0.27217
## Year1998          0.608      0.313   1.94 0.05272 .
## Year1999          0.217      0.326   0.67 0.50598
## Year2000          1.386      0.355   3.90 0.00012 ***
## Year2001         -0.282      0.489  -0.58 0.56474
## Year2002         -0.242      0.443  -0.55 0.58496
## Year2003          0.606      0.281   2.15 0.03195 *
```

```

## Year2004          0.245      0.264      0.93  0.35274
## Year2005          0.159      0.510      0.31  0.75620
## Year2006         -0.243      0.300     -0.81  0.41769
## Year2007         -0.173      0.339     -0.51  0.61070
## Year2008         -0.387      0.339     -1.14  0.25411
## Year2009         -0.954      0.417     -2.29  0.02279 *
## Year2010         -0.747      0.263     -2.85  0.00471 **
## Year2011         -0.180      0.352     -0.51  0.60992
## Year2012         -0.590      0.272     -2.17  0.03100 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.897
## Multiple R-squared:  0.277, Adjusted R-squared:  0.239
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 310 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.157  0.865  0.950  0.903  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.92e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 342"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2907"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9    8   26   36   29   44   49   32   37   38   69   73   71   47   75
## 2011 2012
##   53   37
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    5    5    6    7   12   35   26   34   35   58   62   59   41   60
## 2011 2012
##   47   30
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7    4    5    6    7   11   29   26   30   33   53   57   55   39   57
## 2011 2012
##   46   29
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```

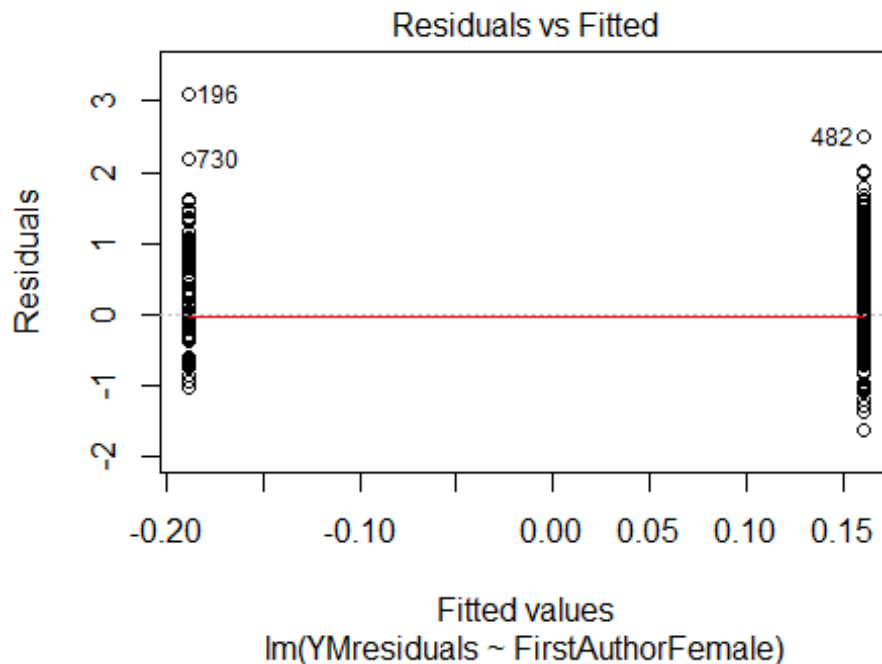


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.1, df = 1, p-value = 0.04
## [1] "Female first author team size 2018 geometric mean: 2.84731468735758"
## [1] "Male first author team size 2018 geometric mean: 2.58581596306162"
```

```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.20310094752884"
## [1] "Male last author team size 2018 geometric mean: 2.48561215189499"

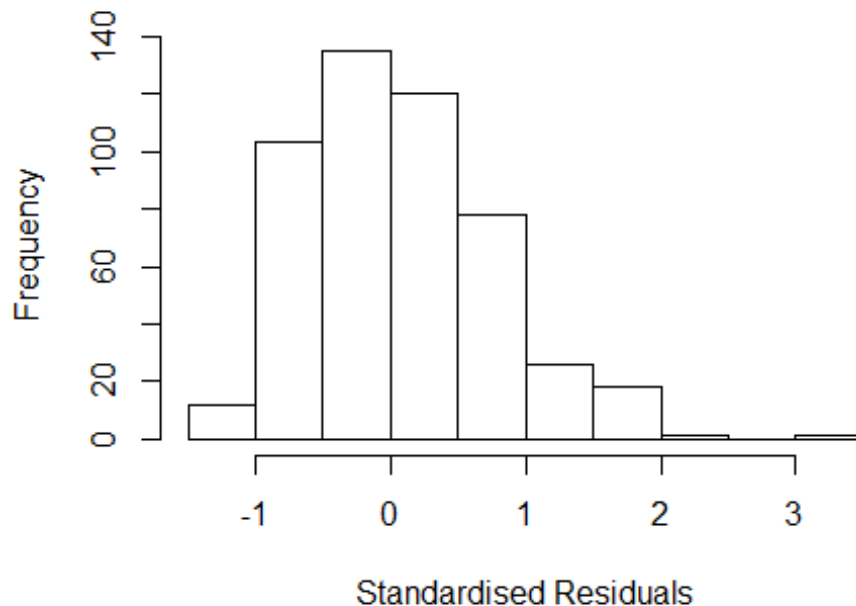
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
##      Year as factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.731 1          1.316
```

## LastAuthorFemale	1.957	1	1.399
## UniqueAuthors	3.352	4	1.163
## Year	3.318	16	1.038

## Residuals from first and last author and team size



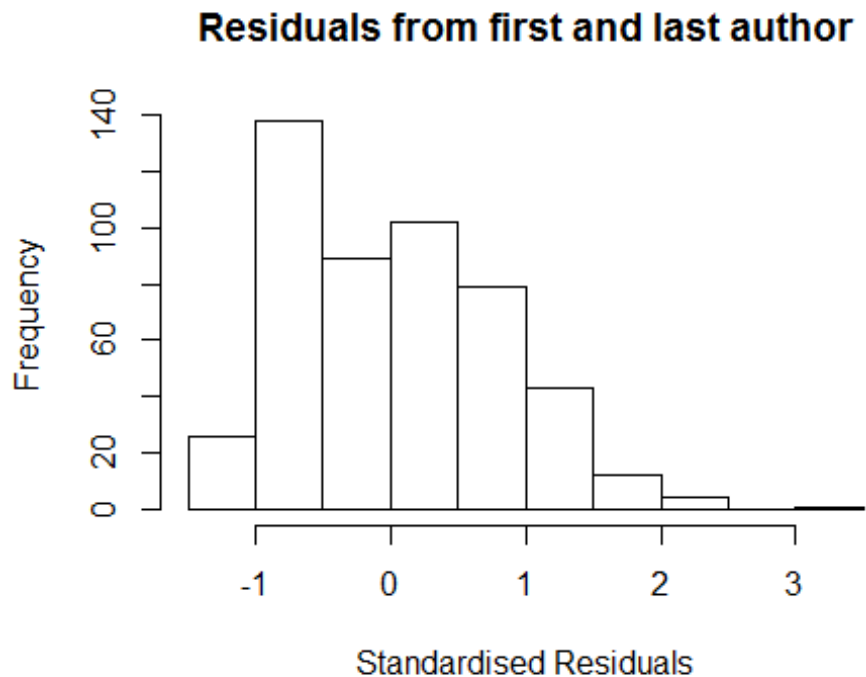
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 196 0036596507 4.111 2002    2907      1    3.253
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.448 -0.471 -0.033  0.500  3.253
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.915266   0.305776   2.99   0.0029 **
## FirstAuthorFemale1 -0.000739   0.078772  -0.01   0.9925
## LastAuthorFemale1 -0.120881   0.084152  -1.44   0.1515
## UniqueAuthors2    0.239557   0.107963   2.22   0.0270 *
## UniqueAuthors3    0.721356   0.124718   5.78  1.3e-08 ***
## UniqueAuthors4    0.908751   0.133773   6.79  3.3e-11 ***
```

```

## UniqueAuthors5      1.115161    0.105703    10.55 < 2e-16 ***
## Year1997             -0.343539    0.491762    -0.70    0.4852
## Year1998             0.416799    0.411867     1.01    0.3121
## Year1999             0.433682    0.403947     1.07    0.2835
## Year2000             0.266965    0.360779     0.74    0.4597
## Year2001            -0.073934    0.371207    -0.20    0.8422
## Year2002             0.064586    0.340584     0.19    0.8497
## Year2003            -0.451036    0.345318    -1.31    0.1921
## Year2004            -0.550462    0.317759    -1.73    0.0839 .
## Year2005            -0.418748    0.314280    -1.33    0.1834
## Year2006            -0.323125    0.320704    -1.01    0.3142
## Year2007            -0.329610    0.322077    -1.02    0.3066
## Year2008            -0.395181    0.309751    -1.28    0.2027
## Year2009            -0.498991    0.318617    -1.57    0.1180
## Year2010            -0.244712    0.319462    -0.77    0.4441
## Year2011            -0.346704    0.320489    -1.08    0.2799
## Year2012            -0.122480    0.319812    -0.38    0.7019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.622
## Multiple R-squared:  0.368, Adjusted R-squared:  0.339
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## observation 57 is an outlier with |weight| = 0 ( < 0.0002);
## 22 weights are ~= 1. The remaining 471 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0813 0.8810 0.9420 0.8970 0.9800 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.270 1          1.507

```

## LastAuthorFemale	2.439	1	1.562
## Year	1.590	16	1.015



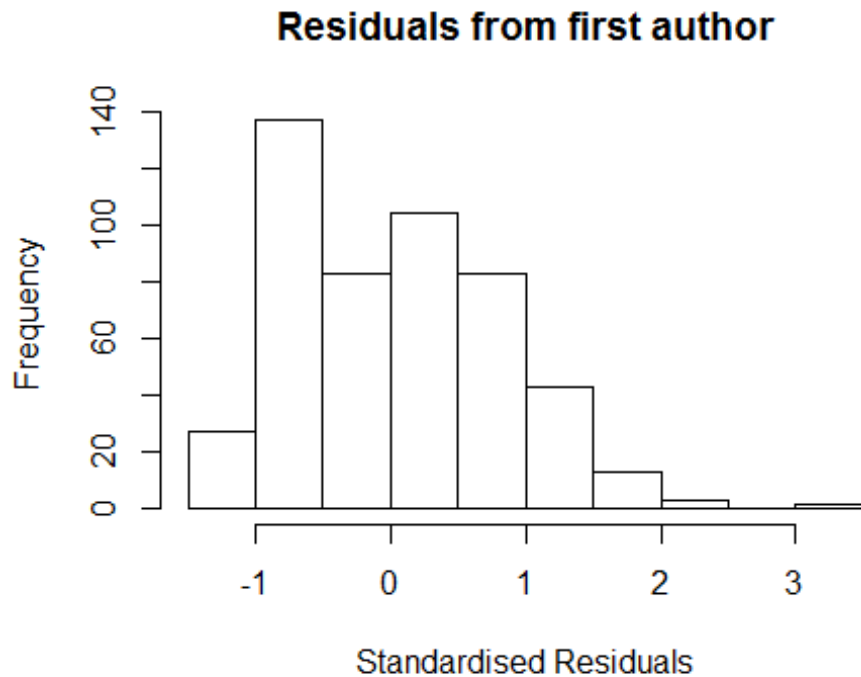
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 196 0036596507 4.111 2002      2907      1      3.124
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4154 -0.6006 -0.0126  0.5748  3.1235
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.516      0.271    5.58   4e-08 ***
## FirstAuthorFemale1 -0.112      0.104   -1.07   0.2855
## LastAuthorFemale1  -0.316      0.105   -3.00   0.0028 **
## Year1997          -0.146      0.495   -0.30   0.7674
## Year1998           0.544      0.297    1.83   0.0676 .
## Year1999           0.380      0.328    1.16   0.2477
## Year2000           0.222      0.311    0.71   0.4755
## Year2001          -0.184      0.302   -0.61   0.5437
## Year2002          -0.100      0.331   -0.30   0.7622
```



```

## Year2003          -0.486      0.313   -1.55   0.1213
## Year2004          -0.721      0.312   -2.31   0.0215 *
## Year2005          -0.455      0.311   -1.46   0.1442
## Year2006          -0.575      0.299   -1.92   0.0549 .
## Year2007          -0.584      0.300   -1.94   0.0526 .
## Year2008          -0.573      0.297   -1.93   0.0547 .
## Year2009          -0.487      0.305   -1.60   0.1111
## Year2010          -0.339      0.304   -1.12   0.2649
## Year2011          -0.455      0.302   -1.50   0.1332
## Year2012          -0.147      0.312   -0.47   0.6371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.802
## Multiple R-squared:  0.145, Adjusted R-squared:  0.112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 39 weights are ~ = 1. The remaining 455 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.096  0.879   0.946   0.915  0.981   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.183 1          1.088
## Year              1.183 16          1.005

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 196 0036596507 4.111 2002      2907      1      3.124
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.49115 -0.59878  0.00118  0.64457  3.06474
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5044     0.2610    5.76 1.5e-08 ***
## FirstAuthorFemale1 -0.3227     0.0773   -4.18 3.5e-05 ***
## Year1997         -0.0132     0.5269   -0.03  0.980
## Year1998          0.5342     0.2906    1.84  0.067 .
## Year1999          0.4656     0.3240    1.44  0.151
## Year2000          0.2493     0.2969    0.84  0.402
## Year2001         -0.1519     0.2935   -0.52  0.605
## Year2002         -0.1354     0.3177   -0.43  0.670
## Year2003         -0.5201     0.3030   -1.72  0.087 .
## Year2004         -0.7716     0.3026   -2.55  0.011 *
## Year2005         -0.5100     0.3051   -1.67  0.095 .
## Year2006         -0.6040     0.2874   -2.10  0.036 *
```

```

## Year2007          -0.5947      0.2912   -2.04    0.042 *
## Year2008          -0.6430      0.2845   -2.26    0.024 *
## Year2009          -0.5829      0.2927   -1.99    0.047 *
## Year2010          -0.3600      0.2951   -1.22    0.223
## Year2011          -0.4769      0.2922   -1.63    0.103
## Year2012          -0.1259      0.3063   -0.41    0.681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.81
## Multiple R-squared:  0.129, Adjusted R-squared:  0.0977
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 444 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.886  0.938  0.912  0.977  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.299 1          1.140
## Year            1.299 16          1.008

## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 196 0036596507 4.111 2002      2907      1      3.124
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.3859 -0.6217 -0.0157  0.6032  3.1179

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.5129    0.2670   5.67 2.5e-08 ***
## LastAuthorFemale1 -0.3927    0.0773  -5.08 5.4e-07 ***
## Year1997          -0.2126    0.4663  -0.46  0.649
## Year1998           0.5400    0.2938   1.84  0.067 .
## Year1999           0.3433    0.3201   1.07  0.284
## Year2000           0.2028    0.3069   0.66  0.509
## Year2001          -0.1920    0.2980  -0.64  0.520
## Year2002          -0.1270    0.3257  -0.39  0.697
## Year2003          -0.4985    0.3108  -1.60  0.109
## Year2004          -0.7341    0.3070  -2.39  0.017 *
## Year2005          -0.4713    0.3048  -1.55  0.123
## Year2006          -0.5890    0.2931  -2.01  0.045 *
## Year2007          -0.5984    0.2945  -2.03  0.043 *
## Year2008          -0.5802    0.2924  -1.98  0.048 *
## Year2009          -0.4883    0.3014  -1.62  0.106
## Year2010          -0.3567    0.2979  -1.20  0.232
## Year2011          -0.4759    0.2963  -1.61  0.109
## Year2012          -0.1895    0.3036  -0.62  0.533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.825
## Multiple R-squared:  0.141, Adjusted R-squared:  0.11
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 441 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.122  0.887   0.945   0.916   0.977   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 494"
## [1] ""

```

```

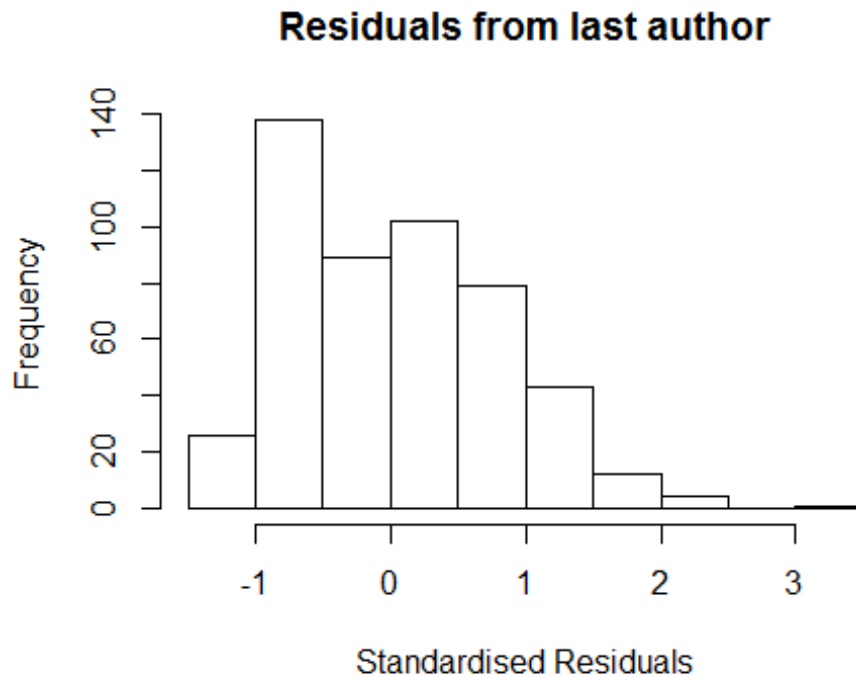
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2908"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2   33   32   41   29   29   31   17   16   14   43   46   31   22   28
## 2011 2012
##   40   35
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    0    0    2    1    2    0    1    1    2   18   25   14   12    9
## 2011 2012
##   20   15
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    0    0    0    2    1    2    0    1    1    2   18   23   14   11    8
## 2011 2012
##   19   14
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.28172829377935"
## [1] "Male first author team size 2018 geometric mean: 3.44787527406753"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 22, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.29277625100892"
## [1] "Male last author team size 2018 geometric mean: 3.22138931910848"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

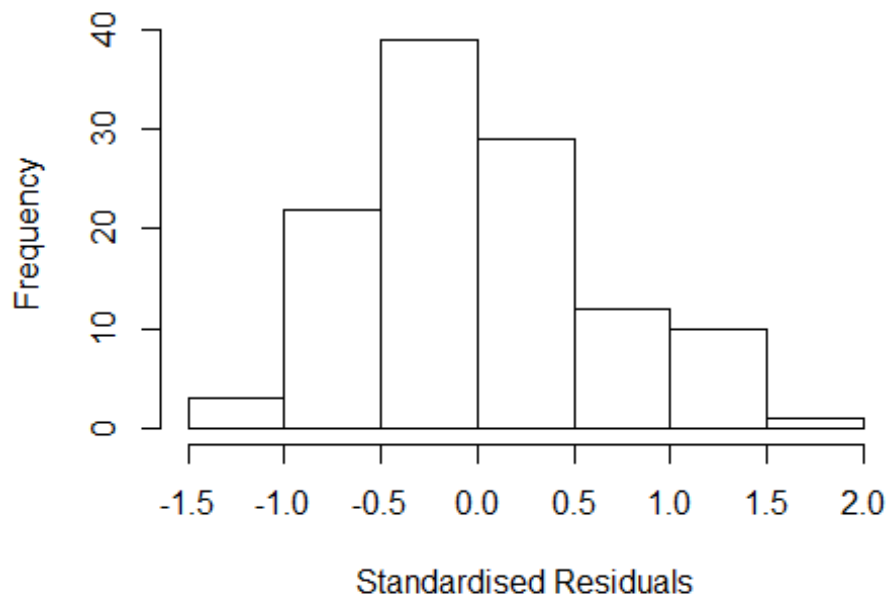
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
## FirstAuthorFemale	3.415e+14	1	1.848e+07
## LastAuthorFemale	-5.067e+13	1	NaN
## UniqueAuthors	9.653e+14	4	7.466e+01
## Year	1.968e+29	12	1.662e+01

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.11218 -0.45778 -0.00985 0.34880 1.60775
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.69141 1.24173 1.36 0.17631
## FirstAuthorFemale1 -0.37219 0.25846 -1.44 0.15309
## LastAuthorFemale1 0.07078 0.26130 0.27 0.78706
## UniqueAuthors2 0.52030 0.16438 3.17 0.00207 **
## UniqueAuthors3 0.29765 0.19897 1.50 0.13790
## UniqueAuthors4 0.07644 0.30644 0.25 0.80355
## UniqueAuthors5 1.18286 0.31217 3.79 0.00026 ***
## Year2000 -0.00352 1.24915 0.00 0.99776
## Year2001 -1.13850 1.23564 -0.92 0.35913
## Year2003 -1.29441 1.24173 -1.04 0.29981
```

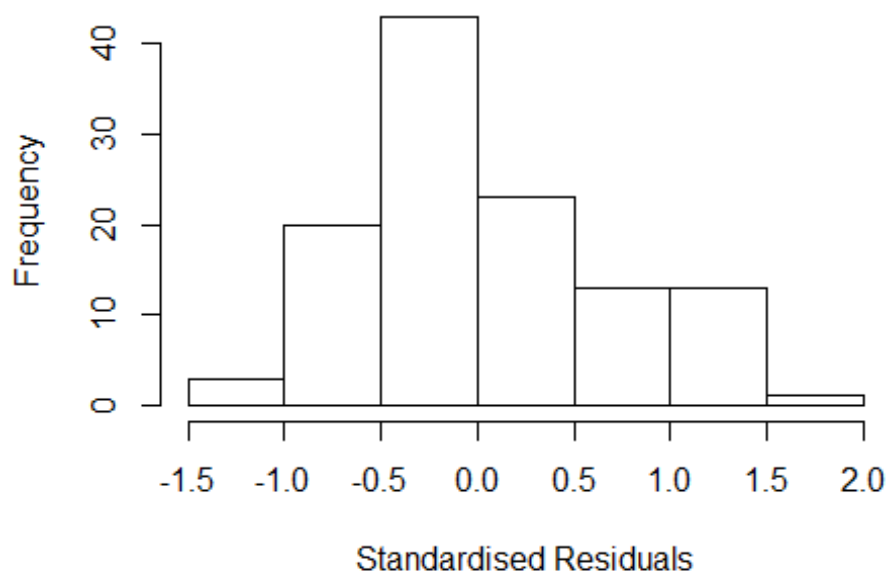
```

## Year2004          -1.01500      1.22186      -0.83      0.40818
## Year2005          -2.10799      1.27472      -1.65      0.10142
## Year2006          -1.16580      1.22969      -0.95      0.34546
## Year2007          -0.87915      1.23500      -0.71      0.47826
## Year2008          -0.57547      1.23934      -0.46      0.64344
## Year2009          -0.50151      1.22989      -0.41      0.68434
## Year2010          -0.36238      1.24997      -0.29      0.77250
## Year2011          -0.61244      1.23395      -0.50      0.62079
## Year2012          -1.19532      1.23127      -0.97      0.33406
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.354, Adjusted R-squared:  0.234
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 105 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.437  0.848   0.945   0.902   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale -7.542e+14  1      NaN
## LastAuthorFemale -4.251e+13  1      NaN
## Year -1.133e+28 12      NaN

```



## Residuals from first and last author



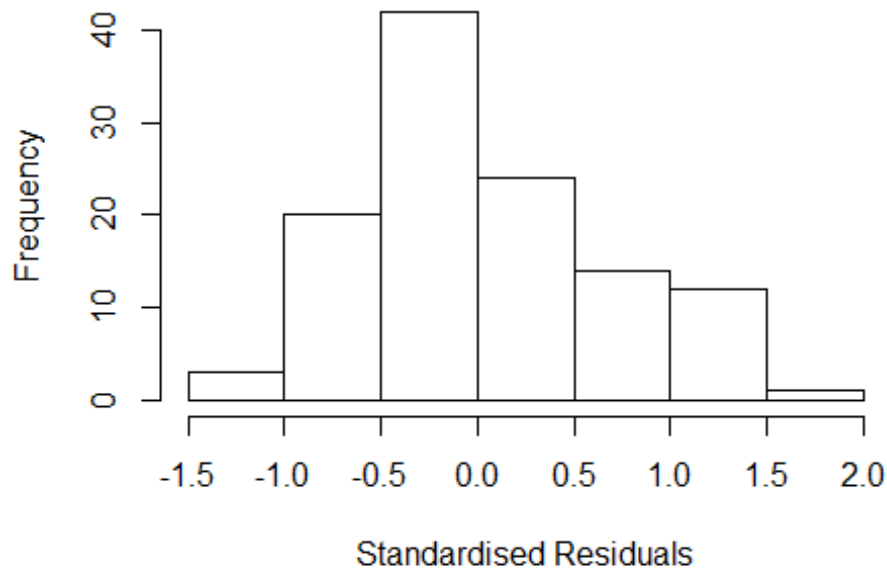
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2678 -0.4104 -0.0375 0.4598 1.7829
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.679 1.165 1.44 0.15
## FirstAuthorFemale1 -0.146 0.302 -0.49 0.63
## LastAuthorFemale1 -0.142 0.275 -0.52 0.61
## Year2000 0.304 1.173 0.26 0.80
## Year2001 -1.138 1.157 -0.98 0.33
## Year2003 -1.282 1.165 -1.10 0.27
## Year2004 -1.015 1.143 -0.89 0.38
## Year2005 -1.362 1.194 -1.14 0.26
## Year2006 -0.980 1.150 -0.85 0.40
## Year2007 -0.602 1.161 -0.52 0.61
## Year2008 -0.122 1.158 -0.11 0.92
## Year2009 -0.295 1.154 -0.26 0.80
```

```

## Year2010          -0.167      1.178   -0.14    0.89
## Year2011          -0.521      1.155   -0.45    0.65
## Year2012          -1.068      1.155   -0.92    0.36
##
## Robust residual standard error: 0.61
## Multiple R-squared:  0.25,   Adjusted R-squared:  0.146
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 98 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.374  0.824  0.936  0.888  0.984  0.999
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00             5.00e-01           4.69e+00           1.00e-07
##           rel.tol             solve.tol           eps.outlier           eps.x
##           1.00e-07             1.00e-07           8.62e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01             5.00e-01
##   nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev      mts    compute.rd
##           0         1000         0
##           psi             subsampling             cov
##           "bisquare"      "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.42e+13 1      5.848e+06
## Year              3.42e+13 12      3.664e+00

```

## Residuals from first author



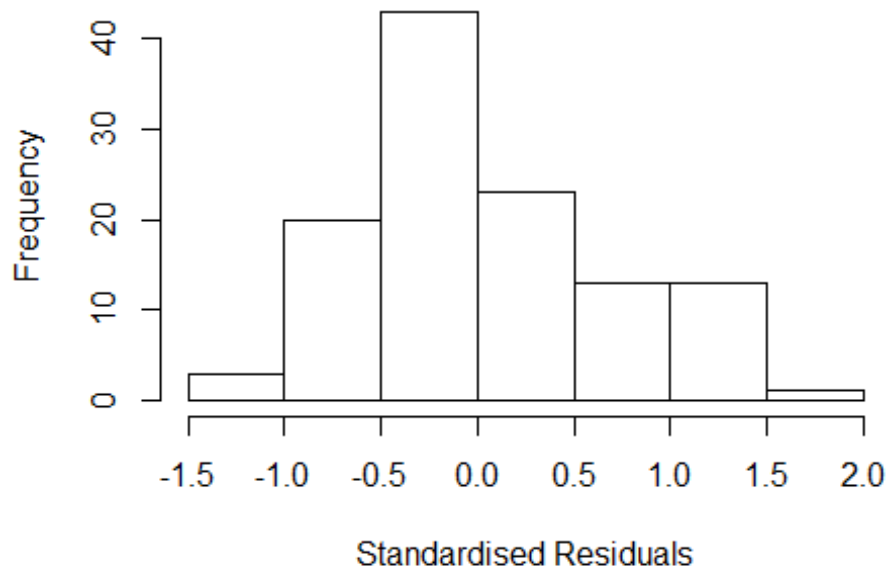
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2525 -0.3753 -0.0469 0.4560 1.6643
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.636 1.197 1.37 0.17
## FirstAuthorFemale1 -0.246 0.198 -1.24 0.22
## Year2000 0.446 1.176 0.38 0.71
## Year2001 -1.138 1.191 -0.96 0.34
## Year2003 -1.239 1.197 -1.04 0.30
## Year2004 -1.015 1.176 -0.86 0.39
## Year2005 -1.461 1.203 -1.21 0.23
## Year2006 -0.966 1.183 -0.82 0.42
## Year2007 -0.583 1.195 -0.49 0.63
## Year2008 -0.138 1.189 -0.12 0.91
## Year2009 -0.241 1.186 -0.20 0.84
## Year2010 -0.163 1.208 -0.13 0.89
```

```

## Year2011          -0.517      1.189   -0.43    0.66
## Year2012          -1.054      1.188   -0.89    0.38
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.248, Adjusted R-squared:  0.152
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 96 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.423  0.817  0.931  0.881  0.977  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale -7.336e+12  1      NaN
## Year              -7.336e+12 12      NaN

```

## Residuals from last author



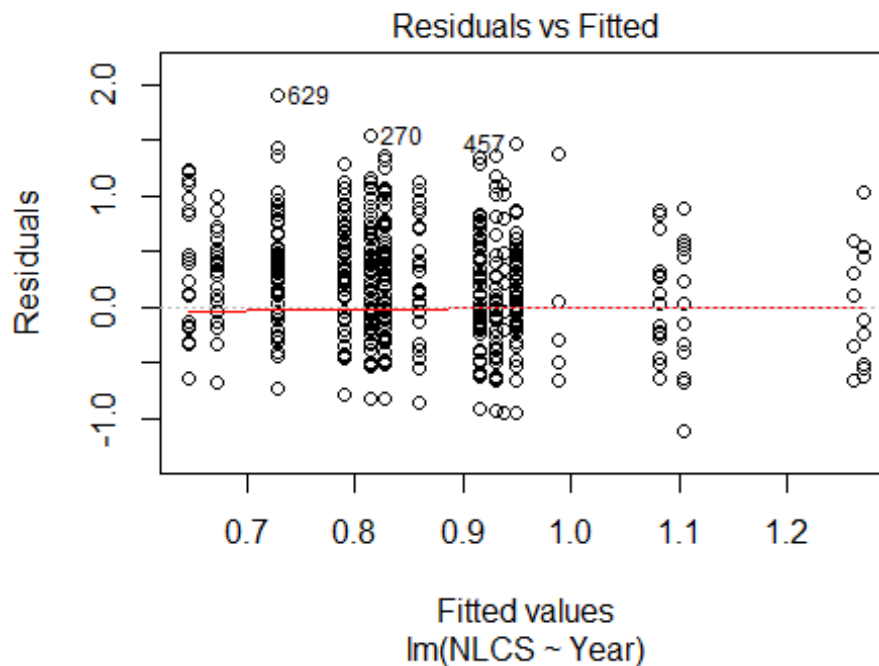
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2990 -0.4142 -0.0522 0.4578 1.9397
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.627 1.192 1.36 0.18
## LastAuthorFemale1 -0.237 0.179 -1.33 0.19
## Year2000 0.209 1.192 0.17 0.86
## Year2001 -1.139 1.192 -0.96 0.34
## Year2003 -1.230 1.192 -1.03 0.30
## Year2004 -1.015 1.177 -0.86 0.39
## Year2005 -1.216 1.184 -1.03 0.31
## Year2006 -0.976 1.184 -0.82 0.41
## Year2007 -0.613 1.194 -0.51 0.61
## Year2008 -0.091 1.189 -0.08 0.94
## Year2009 -0.325 1.187 -0.27 0.79
## Year2010 -0.140 1.212 -0.12 0.91
```

```

## Year2011          -0.519      1.190   -0.44      0.66
## Year2012          -1.051      1.188   -0.88      0.38
##
## Robust residual standard error: 0.6
## Multiple R-squared:  0.251, Adjusted R-squared:  0.156
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 102 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.275  0.817  0.938  0.888  0.984  0.999
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           8.62e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 116"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2909"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   26   28   20   36   60   51   36   35   46   46   70   76  123  104
## 2011 2012
##  119  119
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    5   12    8   13   19   40   26   33   37   37   57   61  110   78
## 2011 2012
##   92   97
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    5   10    8   11   19   35   19   20   29   33   52   55  101   69
## 2011 2012
##   81   90

```

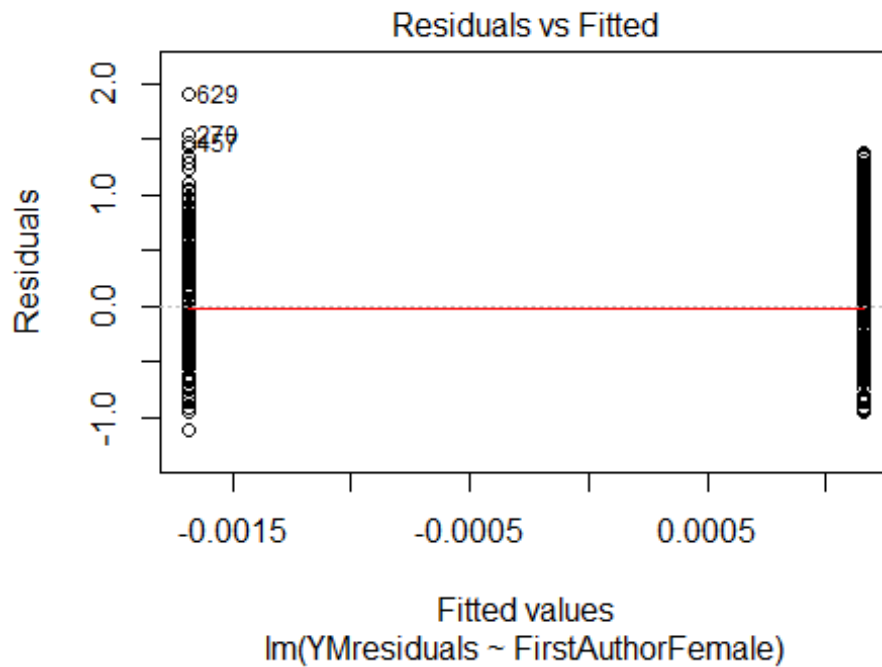
```
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 8.4, df = 16, p-value = 0.9
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 2.96046225953756"
## [1] "Male first author team size 2018 geometric mean: 2.49825154350716"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.6888712067944"
## [1] "Male last author team size 2018 geometric mean: 3.15503004291913"
```

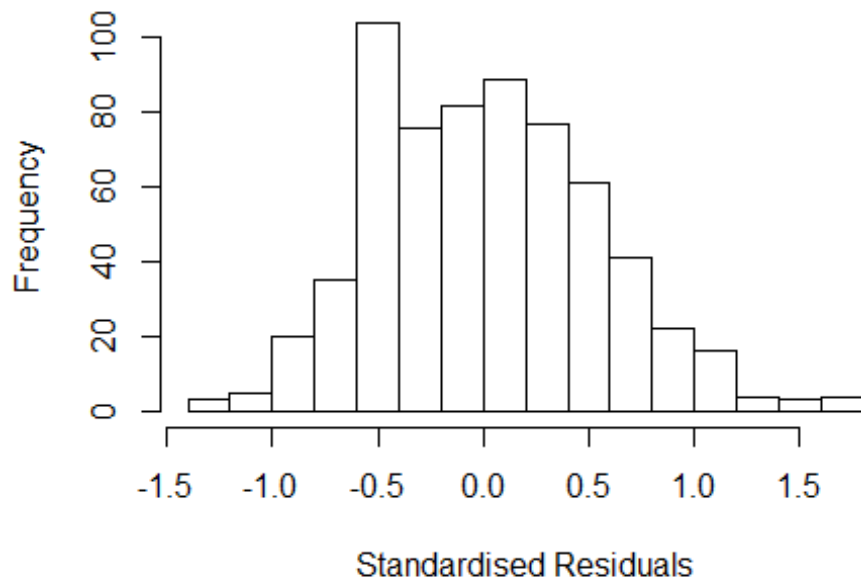
```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 430, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.898 1          1.378
## LastAuthorFemale  2.048 1          1.431
## UniqueAuthors    1.693 4          1.068
## Year              1.730 16         1.017
```



## Residuals from first and last author and team size



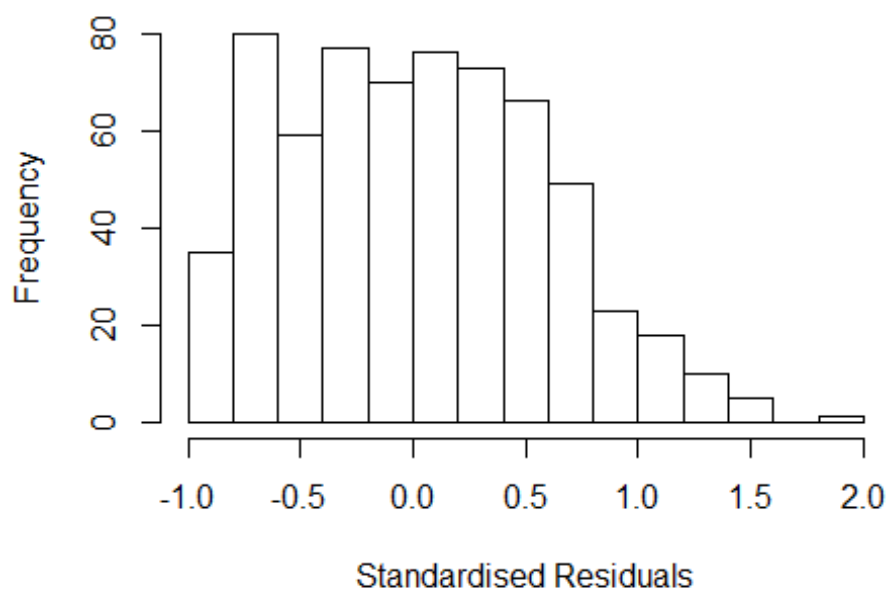
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2260 -0.4018 -0.0043 0.3657 1.7229
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.4126 0.2610 1.58 0.11
## FirstAuthorFemale1 0.0197 0.0574 0.34 0.73
## LastAuthorFemale1 -0.0885 0.0597 -1.48 0.14
## UniqueAuthors2 0.3207 0.0651 4.93 1.1e-06 ***
## UniqueAuthors3 0.4724 0.0721 6.55 1.2e-10 ***
## UniqueAuthors4 0.5109 0.0626 8.16 1.8e-15 ***
## UniqueAuthors5 0.5976 0.0778 7.68 6.3e-14 ***
## Year1997 0.5513 0.2811 1.96 0.05 .
## Year1998 0.0836 0.3318 0.25 0.80
## Year1999 0.6011 0.3793 1.59 0.11
```

```

## Year2000          0.2485      0.3043      0.82      0.41
## Year2001          0.3447      0.2777      1.24      0.21
## Year2002          0.2019      0.2711      0.74      0.46
## Year2003          0.0731      0.2870      0.25      0.80
## Year2004         -0.1556      0.2759     -0.56      0.57
## Year2005         -0.0752      0.2835     -0.27      0.79
## Year2006          0.1861      0.2811      0.66      0.51
## Year2007          0.2845      0.2695      1.06      0.29
## Year2008          0.1476      0.2673      0.55      0.58
## Year2009          0.0579      0.2635      0.22      0.83
## Year2010          0.2512      0.2639      0.95      0.34
## Year2011          0.1750      0.2666      0.66      0.51
## Year2012          0.1051      0.2645      0.40      0.69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.221, Adjusted R-squared:  0.193
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 597 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.287  0.881  0.947  0.912  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.56e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.865 1      1.366
## LastAuthorFemale  1.895 1      1.377
## Year              1.202 16      1.006

```

## Residuals from first and last author



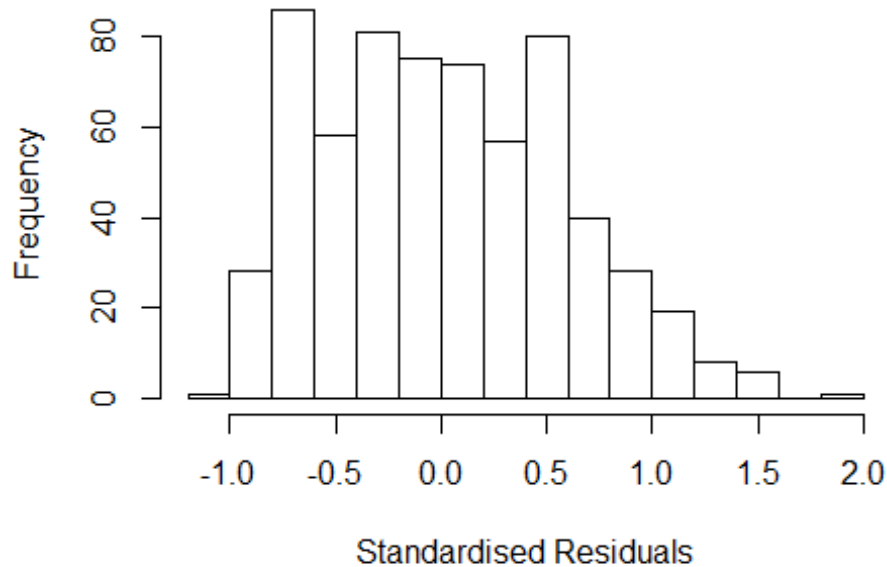
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.988453 -0.448687 0.000832 0.439569 1.920556
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8652 0.2918 2.96 0.0031 **
## FirstAuthorFemale1 0.0830 0.0642 1.29 0.1967
## LastAuthorFemale1 -0.1777 0.0642 -2.77 0.0058 **
## Year1997 0.4423 0.3608 1.23 0.2207
## Year1998 -0.0884 0.3995 -0.22 0.8250
## Year1999 0.3650 0.3645 1.00 0.3170
## Year2000 0.1233 0.3466 0.36 0.7222
## Year2001 0.2276 0.3097 0.73 0.4628
## Year2002 0.0987 0.3093 0.32 0.7498
## Year2003 -0.0662 0.3152 -0.21 0.8338
## Year2004 -0.4158 0.3069 -1.35 0.1760
## Year2005 -0.2700 0.3152 -0.86 0.3919
```

```

## Year2006          0.0118      0.3154      0.04      0.9701
## Year2007          0.1421      0.3027      0.47      0.6389
## Year2008         -0.0337      0.3050     -0.11      0.9119
## Year2009         -0.1547      0.2982     -0.52      0.6040
## Year2010          0.0860      0.3005      0.29      0.7749
## Year2011         -0.0201      0.3016     -0.07      0.9470
## Year2012         -0.0666      0.3003     -0.22      0.8245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.0744, Adjusted R-squared:  0.0477
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 56 weights are ~= 1. The remaining 586 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.329  0.887   0.947   0.921   0.983   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.56e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.11 1          1.054
## Year              1.11 16          1.003

```

## Residuals from first author



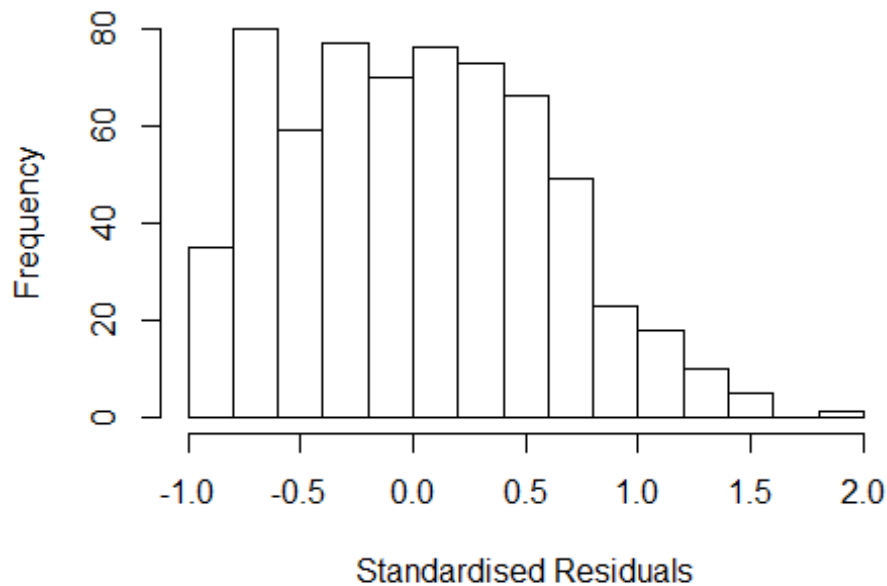
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.01231 -0.45736 -0.00863 0.44747 1.95850
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8710 0.3054 2.85 0.0045 **
## FirstAuthorFemale1 -0.0131 0.0498 -0.26 0.7929
## Year1997 0.4002 0.3748 1.07 0.2860
## Year1998 -0.1121 0.4141 -0.27 0.7868
## Year1999 0.3750 0.3744 1.00 0.3170
## Year2000 0.1413 0.3612 0.39 0.6958
## Year2001 0.2036 0.3234 0.63 0.5293
## Year2002 0.0514 0.3234 0.16 0.8736
## Year2003 -0.1452 0.3300 -0.44 0.6602
## Year2004 -0.4819 0.3208 -1.50 0.1336
## Year2005 -0.3136 0.3270 -0.96 0.3379
## Year2006 -0.0284 0.3281 -0.09 0.9310
```

```

## Year2007          0.0945      0.3173      0.30      0.7660
## Year2008         -0.0734      0.3187     -0.23      0.8180
## Year2009         -0.1985      0.3124     -0.64      0.5254
## Year2010          0.0387      0.3146      0.12      0.9020
## Year2011         -0.0601      0.3150     -0.19      0.8488
## Year2012         -0.1131      0.3144     -0.36      0.7191
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.635
## Multiple R-squared:  0.0609, Adjusted R-squared:  0.0353
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 61 weights are ~= 1. The remaining 581 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.321  0.879   0.946   0.921   0.981   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.56e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.12  1      1.058
## Year              1.12 16      1.004

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0090 -0.4510 -0.0046 0.4402 1.8981
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8936 0.2926 3.05 0.0024 **
## LastAuthorFemale1 -0.1331 0.0498 -2.67 0.0077 **
## Year1997 0.4309 0.3601 1.20 0.2320
## Year1998 -0.1059 0.4031 -0.26 0.7928
## Year1999 0.3496 0.3645 0.96 0.3378
## Year2000 0.1154 0.3496 0.33 0.7414
## Year2001 0.2162 0.3095 0.70 0.4851
## Year2002 0.0903 0.3094 0.29 0.7705
## Year2003 -0.0689 0.3146 -0.22 0.8267
## Year2004 -0.4226 0.3064 -1.38 0.1683
## Year2005 -0.2630 0.3127 -0.84 0.4007
## Year2006 0.0102 0.3145 0.03 0.9740
```

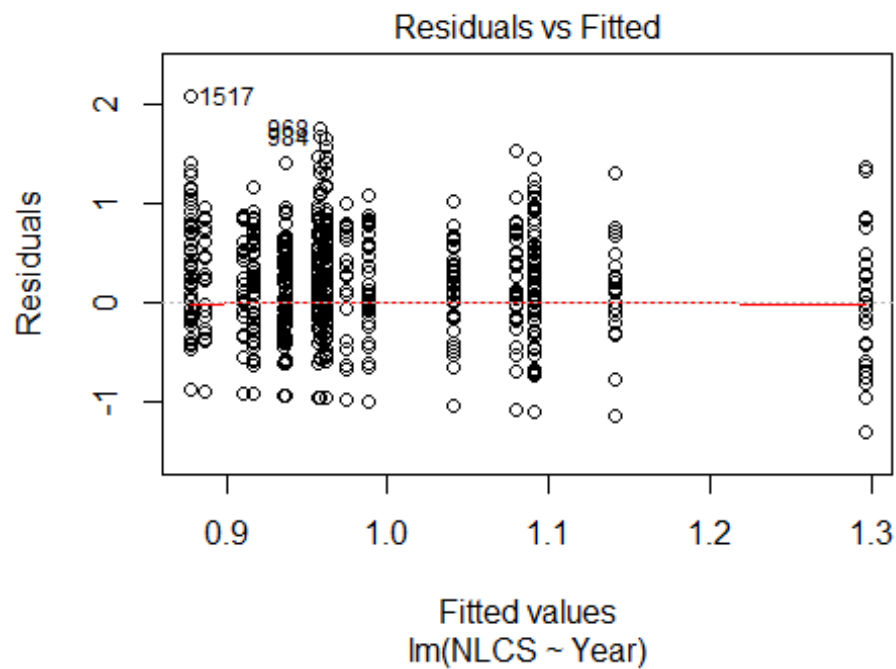
```

## Year2007          0.1402      0.3029      0.46      0.6435
## Year2008          -0.0410      0.3049     -0.13      0.8930
## Year2009          -0.1607      0.2983     -0.54      0.5902
## Year2010           0.0839      0.3005      0.28      0.7801
## Year2011          -0.0199      0.3014     -0.07      0.9474
## Year2012          -0.0654      0.3000     -0.22      0.8274
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.0713, Adjusted R-squared:  0.046
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 580 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.341  0.883   0.945   0.920   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.56e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 642"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2910"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   63   84   67   56   53   70   57   42   42   51   95  114  118  125  138
## 2011 2012
##  156  122
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   47   52   48   45   32   27   31   26   28   29   66   74   83   74  100
## 2011 2012

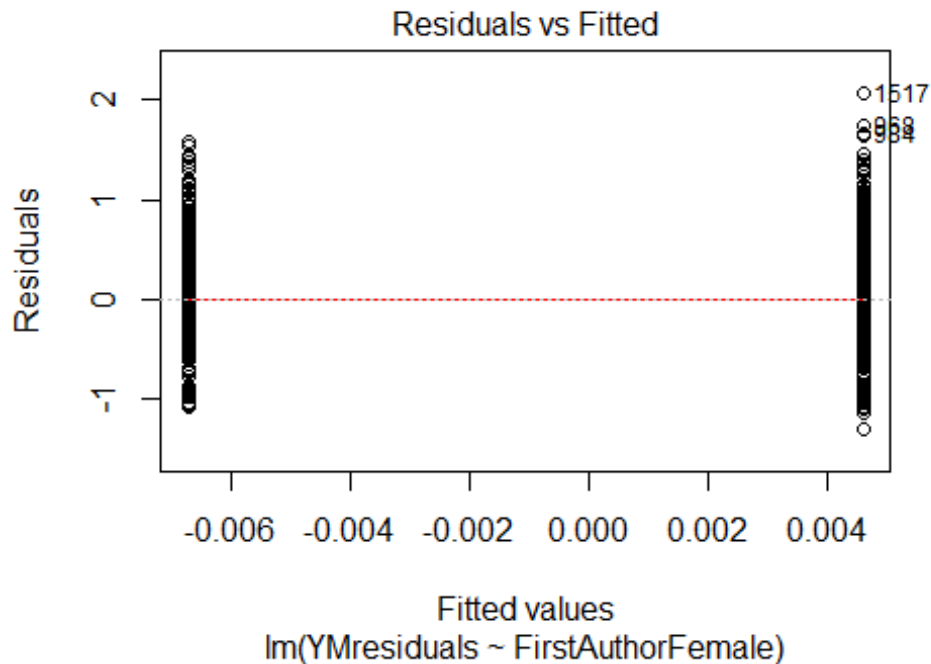
```



```
## 133 101
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 46 50 47 45 31 26 29 21 24 22 55 71 79 70 97
## 2011 2012
## 128 98
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```

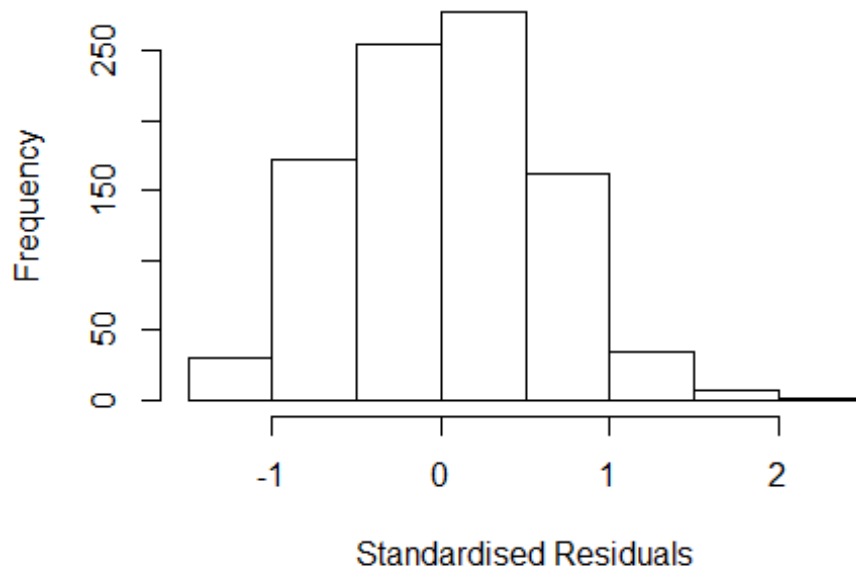


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.6, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 1.88225960830809"
## [1] "Male first author team size 2018 geometric mean: 1.39217787027601"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.8559531939845"
## [1] "Male last author team size 2018 geometric mean: 1.43536265894748"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.913 1      1.383
## LastAuthorFemale  1.923 1      1.387
## UniqueAuthors    1.420 4      1.045
## Year              1.543 16     1.014
```

## Residuals from first and last author and team size



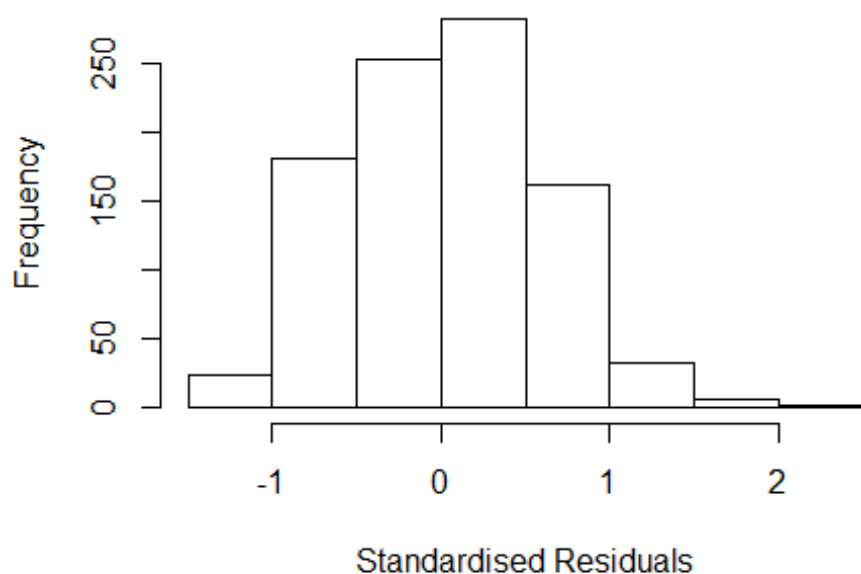
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2552 -0.4022 0.0123 0.4288 2.1498
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05265 0.10604 9.93 <2e-16 ***
## FirstAuthorFemale1 0.00109 0.05826 0.02 0.9851
## LastAuthorFemale1 -0.02147 0.05753 -0.37 0.7091
## UniqueAuthors2 0.05643 0.05807 0.97 0.3314
## UniqueAuthors3 0.20142 0.07343 2.74 0.0062 **
## UniqueAuthors4 0.18909 0.10605 1.78 0.0749 .
## UniqueAuthors5 0.09841 0.09495 1.04 0.3003
## Year1997 -0.08580 0.13198 -0.65 0.5158
## Year1998 -0.11350 0.12262 -0.93 0.3549
## Year1999 -0.03788 0.12813 -0.30 0.7676
```

```

## Year2000      -0.13955    0.13482   -1.04    0.3009
## Year2001      0.04277    0.15270    0.28    0.7795
## Year2002     -0.16089    0.15501   -1.04    0.2996
## Year2003     -0.21714    0.18645   -1.16    0.2445
## Year2004      0.13292    0.18582    0.72    0.4746
## Year2005     -0.24753    0.21966   -1.13    0.2601
## Year2006     -0.26645    0.12445   -2.14    0.0325 *
## Year2007     -0.11782    0.12427   -0.95    0.3433
## Year2008     -0.13689    0.12724   -1.08    0.2823
## Year2009     -0.14140    0.13597   -1.04    0.2987
## Year2010     -0.00239    0.12337   -0.02    0.9846
## Year2011     -0.13355    0.12241   -1.09    0.2756
## Year2012     -0.25449    0.12352   -2.06    0.0396 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0339, Adjusted R-squared:  0.0107
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 854 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.185  0.854  0.945  0.909  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.905 1      1.380
## LastAuthorFemale  1.897 1      1.377
## Year              1.103 16      1.003

```

## Residuals from first and last author



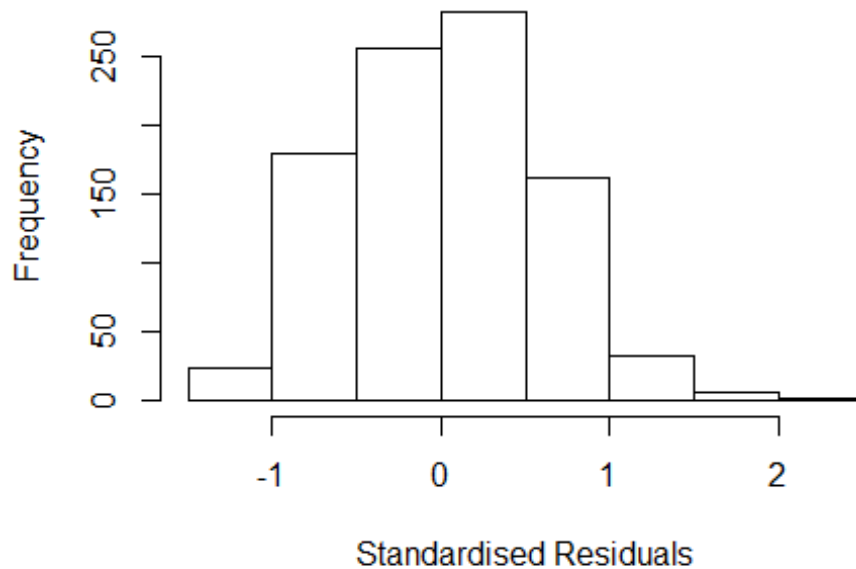
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2261 -0.4233 0.0159 0.4320 2.1104
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0774 0.1038 10.38 <2e-16 ***
## FirstAuthorFemale1 0.0106 0.0588 0.18 0.856
## LastAuthorFemale1 -0.0162 0.0578 -0.28 0.779
## Year1997 -0.0946 0.1320 -0.72 0.474
## Year1998 -0.1337 0.1211 -1.10 0.270
## Year1999 -0.0257 0.1266 -0.20 0.839
## Year2000 -0.1122 0.1341 -0.84 0.403
## Year2001 0.0827 0.1500 0.55 0.582
## Year2002 -0.1493 0.1516 -0.99 0.325
## Year2003 -0.1929 0.1838 -1.05 0.294
## Year2004 0.1487 0.1903 0.78 0.435
## Year2005 -0.2329 0.2084 -1.12 0.264
```

```

## Year2006          -0.2620      0.1229   -2.13    0.033 *
## Year2007          -0.1164      0.1230   -0.95    0.344
## Year2008          -0.1345      0.1264   -1.06    0.288
## Year2009          -0.1520      0.1355   -1.12    0.262
## Year2010           0.0132      0.1218    0.11    0.913
## Year2011          -0.1213      0.1215   -1.00    0.319
## Year2012          -0.2398      0.1223   -1.96    0.050 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0238, Adjusted R-squared:  0.00472
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 858 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##   0.213  0.850  0.947   0.910   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##           "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1          1.021
## Year              1.042 16          1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.225 -0.426 0.012 0.434 2.114
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.075545 0.103567 10.39 <2e-16 ***
## FirstAuthorFemale1 -0.000151 0.043511 0.00 0.997
## Year1997 -0.094644 0.132050 -0.72 0.474
## Year1998 -0.134763 0.121230 -1.11 0.267
## Year1999 -0.024638 0.126517 -0.19 0.846
## Year2000 -0.112212 0.134211 -0.84 0.403
## Year2001 0.079998 0.149444 0.54 0.593
## Year2002 -0.148332 0.151312 -0.98 0.327
## Year2003 -0.191242 0.183980 -1.04 0.299
## Year2004 0.149184 0.190321 0.78 0.433
## Year2005 -0.231794 0.207860 -1.12 0.265
## Year2006 -0.262276 0.122990 -2.13 0.033 *
```

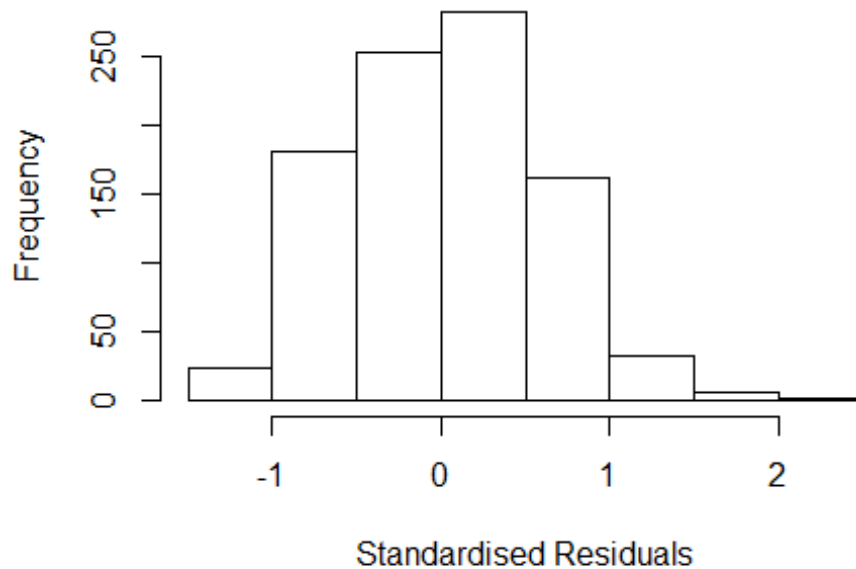
```

## Year2007          -0.115763    0.122987    -0.94    0.347
## Year2008          -0.135748    0.126364    -1.07    0.283
## Year2009          -0.151447    0.135530    -1.12    0.264
## Year2010           0.013045    0.121905     0.11    0.915
## Year2011          -0.120526    0.121590    -0.99    0.322
## Year2012          -0.241284    0.122350    -1.97    0.049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.00567
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 82 weights are ~= 1. The remaining 857 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.213  0.851  0.947  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.037 1          1.018
## Year             1.037 16          1.001

```



## Residuals from last author



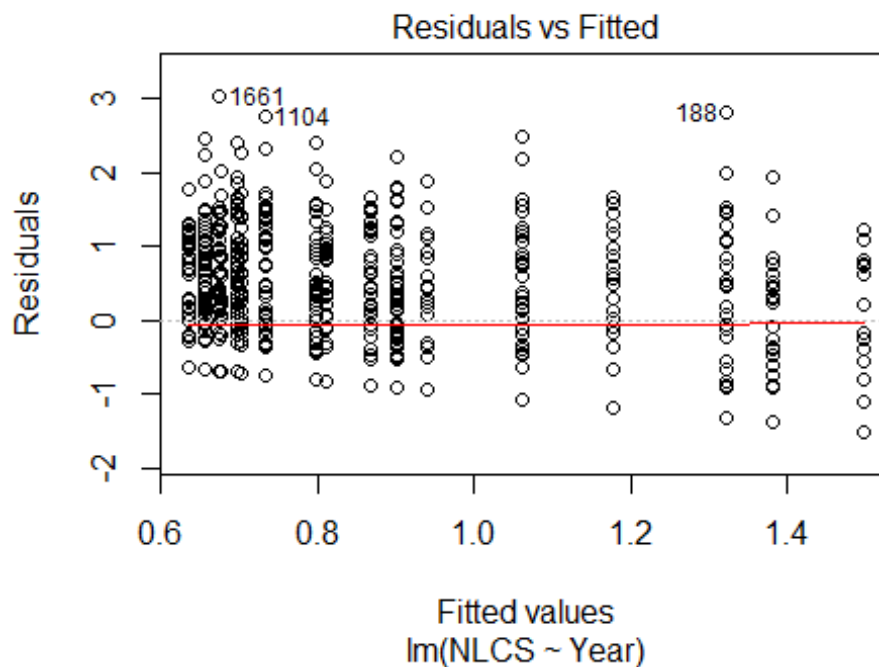
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2278 -0.4214 0.0171 0.4304 2.1096
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.07858 0.10391 10.38 <2e-16 ***
## LastAuthorFemale1 -0.00907 0.04268 -0.21 0.832
## Year1997 -0.09484 0.13215 -0.72 0.473
## Year1998 -0.13385 0.12133 -1.10 0.270
## Year1999 -0.02466 0.12650 -0.19 0.845
## Year2000 -0.11156 0.13419 -0.83 0.406
## Year2001 0.08106 0.14958 0.54 0.588
## Year2002 -0.14904 0.15159 -0.98 0.326
## Year2003 -0.19240 0.18404 -1.05 0.296
## Year2004 0.14927 0.19030 0.78 0.433
## Year2005 -0.23129 0.20779 -1.11 0.266
## Year2006 -0.26158 0.12304 -2.13 0.034 *
```

```

## Year2007          -0.11540      0.12299    -0.94      0.348
## Year2008          -0.13452      0.12642    -1.06      0.288
## Year2009          -0.15114      0.13556    -1.11      0.265
## Year2010           0.01356      0.12198      0.11      0.912
## Year2011          -0.12040      0.12171    -0.99      0.323
## Year2012          -0.24016      0.12233    -1.96      0.050 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0237, Adjusted R-squared:  0.00568
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 83 weights are ~= 1. The remaining 856 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.218  0.851  0.947  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.06e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 939"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2911"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   49  110  108   97  134  127  112   88   71   92  100  126  131  104   86
## 2011 2012
##   84   76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   36   36   27   30   33   83   69   61   82   87  112  120   96   76
## 2011 2012

```

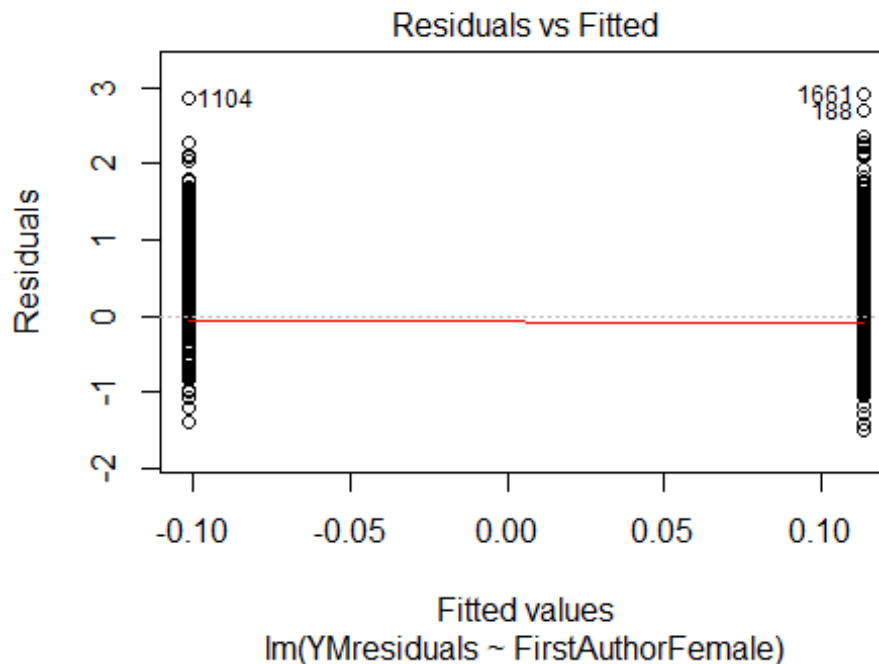
```
## 68 65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 36 33 27 28 29 75 67 60 77 79 108 118 96 74
## 2011 2012
## 65 62
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 26, df = 16, p-value = 0.05
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.8, df = 1, p-value = 0.005
## [1] "Female first author team size 2018 geometric mean: 1.75674804352648"
## [1] "Male first author team size 2018 geometric mean: 1.80300842896086"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

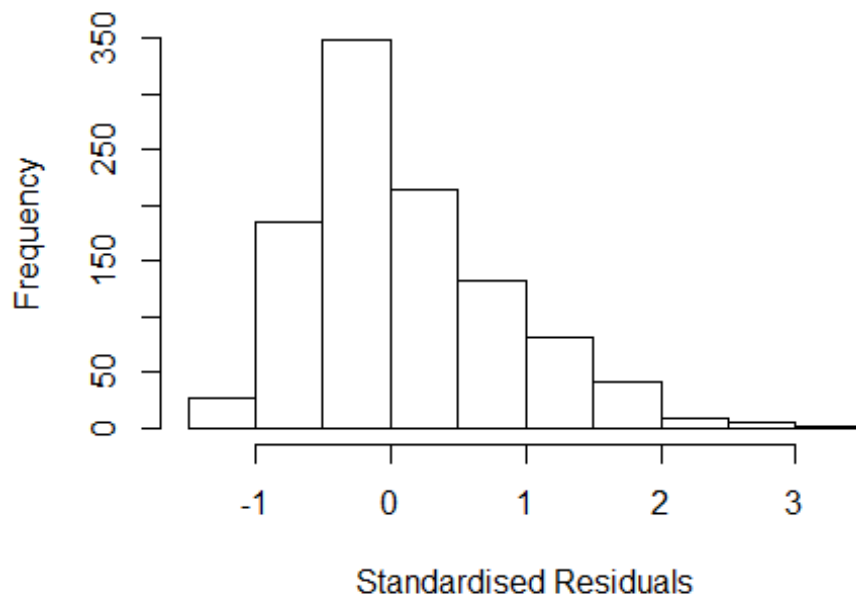
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.12774215707801"
## [1] "Male last author team size 2018 geometric mean: 1.61917861186139"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 240, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.604 1      1.266
## LastAuthorFemale  1.636 1      1.279
## UniqueAuthors    1.582 4      1.059
## Year              1.920 16     1.021
```

## Residuals from first and last author and team size



```
## [1] "List of 7 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 188   0031664637 4.128 1998    2739      3      2.928
## 655   0036651445 2.969 2002    2911      1      2.563
## 874   9444220846 3.535 2004    2911      1      2.634
## 1104  33644878802 3.504 2006    2911      1      3.211
## 1193  34247464253 3.124 2007    2911      1      2.644
## 1322  42149147470 3.098 2008    2911      1      2.534
## 1661  84864553931 3.702 2011    2714      3      2.604
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4547 -0.4670 -0.0646  0.5185  3.2109
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1175     0.2668   4.19 3.1e-05 ***
## FirstAuthorFemale1  0.0530     0.0568   0.93  0.3514
## LastAuthorFemale1  0.1225     0.0586   2.09  0.0367 *
## UniqueAuthors2     0.2083     0.0632   3.30  0.0010 **
```

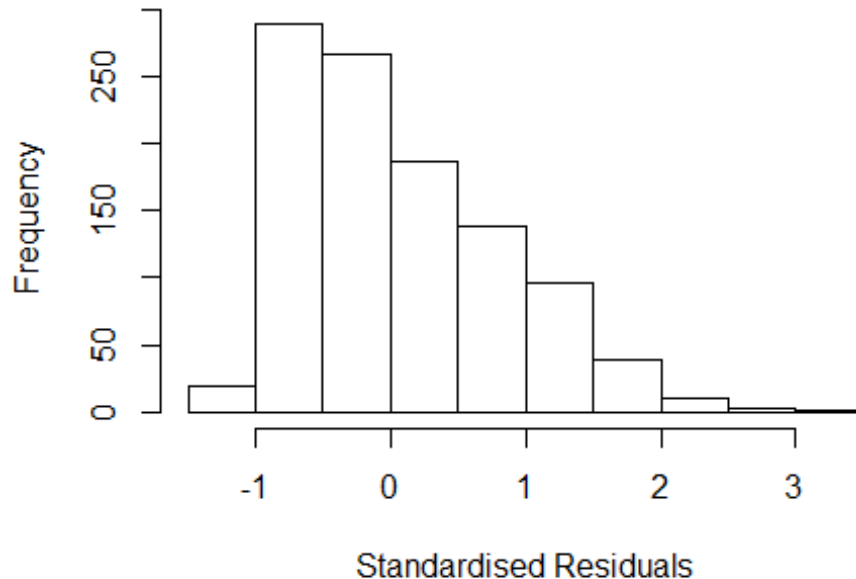
```

## UniqueAuthors3      0.4317      0.0774      5.58 3.2e-08 ***
## UniqueAuthors4      0.6125      0.0901      6.80 1.8e-11 ***
## UniqueAuthors5      0.7891      0.0970      8.13 1.2e-15 ***
## Year1997             -0.7821      0.2815     -2.78 0.0056 **
## Year1998             -0.4021      0.3550     -1.13 0.2576
## Year1999             -0.3438      0.3447     -1.00 0.3187
## Year2000             -0.1361      0.2915     -0.47 0.6407
## Year2001             -0.6603      0.3069     -2.15 0.0317 *
## Year2002             -0.8872      0.2694     -3.29 0.0010 **
## Year2003             -0.7001      0.2730     -2.56 0.0105 *
## Year2004             -0.6002      0.2830     -2.12 0.0342 *
## Year2005             -0.7031      0.2718     -2.59 0.0098 **
## Year2006             -0.8244      0.2769     -2.98 0.0030 **
## Year2007             -0.8132      0.2682     -3.03 0.0025 **
## Year2008             -0.7288      0.2733     -2.67 0.0078 **
## Year2009             -0.6273      0.2732     -2.30 0.0218 *
## Year2010             -0.8213      0.2753     -2.98 0.0029 **
## Year2011             -0.8075      0.2707     -2.98 0.0029 **
## Year2012             -0.6098      0.2823     -2.16 0.0310 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.196, Adjusted R-squared:  0.179
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 2 observations c(61,501) are outliers with |weight| = 0 ( < 9.5e-05);
## 65 weights are ~= 1. The remaining 982 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0263 0.8500 0.9430 0.8770 0.9780 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))

```

```
## FirstAuthorFemale 1.449 1 1.204
## LastAuthorFemale 1.418 1 1.191
## Year 1.337 16 1.009
```

### Residuals from first and last author



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 188   0031664637 4.128 1998    2739      3      2.964
## 874   9444220846 3.535 2004    2911      1      2.504
## 1104 33644878802 3.504 2006    2911      1      2.956
## 1661 84864553931 3.702 2011    2714      3      3.001
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4551 -0.5248 -0.0902  0.5940  3.0007
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.3806    0.2662   5.19 2.6e-07 ***
## FirstAuthorFemale1 0.1692    0.0593   2.85 0.00441 **
## LastAuthorFemale1 0.0501    0.0613   0.82 0.41388
## Year1997       -0.8558    0.2939  -2.91 0.00367 **
## Year1998       -0.3857    0.3495  -1.10 0.27003
```

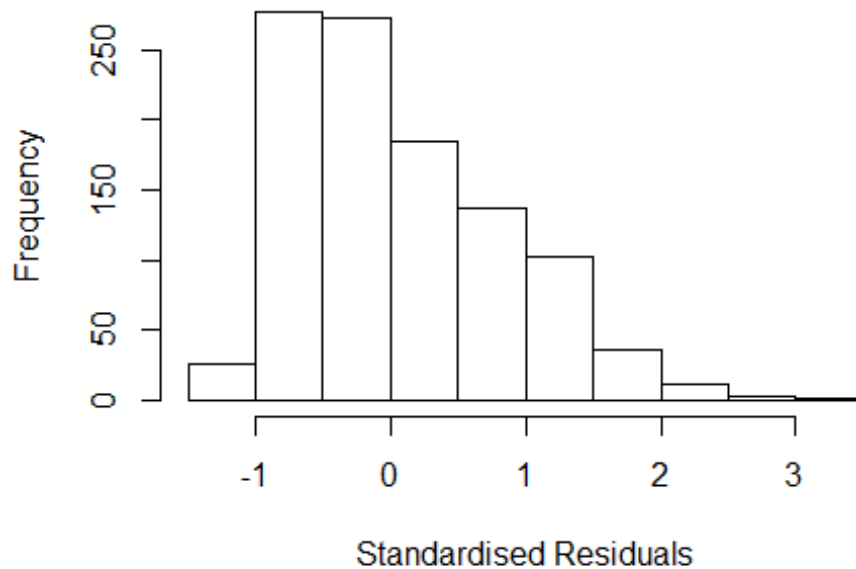
```

## Year1999          -0.3001      0.3742    -0.80    0.42266
## Year2000          -0.1448      0.3060    -0.47    0.63623
## Year2001          -0.6840      0.2993    -2.29    0.02249 *
## Year2002          -0.8516      0.2724    -3.13    0.00182 **
## Year2003          -0.6812      0.2786    -2.45    0.01464 *
## Year2004          -0.5687      0.2911    -1.95    0.05104 .
## Year2005          -0.7770      0.2748    -2.83    0.00477 **
## Year2006          -0.8323      0.2826    -2.95    0.00330 **
## Year2007          -0.8977      0.2733    -3.29    0.00105 **
## Year2008          -0.8574      0.2740    -3.13    0.00180 **
## Year2009          -0.6846      0.2756    -2.48    0.01315 *
## Year2010          -0.9217      0.2756    -3.34    0.00086 ***
## Year2011          -0.8986      0.2790    -3.22    0.00132 **
## Year2012          -0.6647      0.2829    -2.35    0.01900 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.7
## Multiple R-squared:  0.0826, Adjusted R-squared:  0.0666
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 994 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0263 0.8720 0.9450 0.8930 0.9750 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.53e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.169 1          1.081
## Year              1.169 16          1.005

```



## Residuals from first author



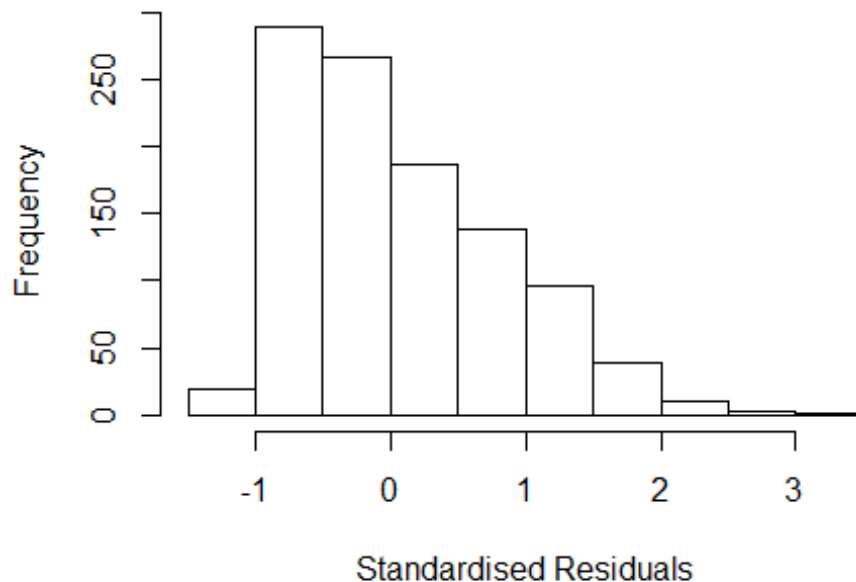
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 188   0031664637 4.128 1998    2739      3    2.964
## 874   9444220846 3.535 2004    2911      1    2.504
## 1104 33644878802 3.504 2006    2911      1    2.956
## 1661 84864553931 3.702 2011    2714      3    3.001
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4258 -0.5301 -0.0888  0.6001  3.0165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.3856    0.2638     5.25 1.8e-07 ***
## FirstAuthorFemale1  0.1930    0.0533     3.62 0.00031 ***
## Year1997       -0.8630    0.2912    -2.96 0.00311 **
## Year1998       -0.3829    0.3494    -1.10 0.27347
## Year1999       -0.3037    0.3744    -0.81 0.41743
## Year2000       -0.1528    0.3023    -0.51 0.61344
## Year2001       -0.6835    0.2973    -2.30 0.02171 *
## Year2002       -0.8451    0.2708    -3.12 0.00185 **
## Year2003       -0.6700    0.2768    -2.42 0.01565 *
```

```

## Year2004          -0.5610      0.2897    -1.94  0.05308 .
## Year2005          -0.7700      0.2729    -2.82  0.00487 **
## Year2006          -0.8294      0.2810    -2.95  0.00323 **
## Year2007          -0.8968      0.2715    -3.30  0.00099 ***
## Year2008          -0.8555      0.2722    -3.14  0.00172 **
## Year2009          -0.6837      0.2736    -2.50  0.01260 *
## Year2010          -0.9170      0.2737    -3.35  0.00084 ***
## Year2011          -0.8931      0.2771    -3.22  0.00131 **
## Year2012          -0.6585      0.2808    -2.35  0.01919 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.699
## Multiple R-squared:  0.0819, Adjusted R-squared:  0.0667
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 55 weights are ~= 1. The remaining 994 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0227 0.8720 0.9430 0.8920 0.9760 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.141 1      1.068
## Year      1.141 16      1.004

```

## Residuals from last author



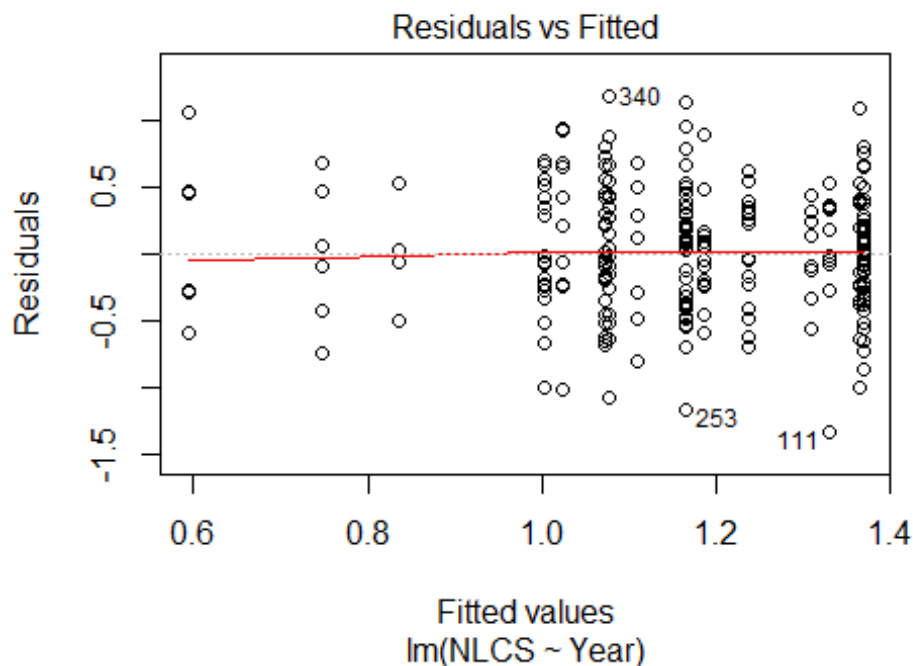
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 188   0031664637 4.128 1998    2739      3    2.964
## 874   9444220846 3.535 2004    2911      1    2.504
## 1104 33644878802 3.504 2006    2911      1    2.956
## 1661 84864553931 3.702 2011    2714      3    3.001
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.457 -0.555 -0.113  0.601  3.083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4563    0.2691   5.41 7.8e-08 ***
## LastAuthorFemale1  0.1329    0.0549   2.42 0.01576 *
## Year1997         -0.8928    0.3009  -2.97 0.00307 **
## Year1998         -0.4114    0.3569  -1.15 0.24924
## Year1999         -0.3339    0.3787  -0.88 0.37817
## Year2000         -0.1319    0.3081  -0.43 0.66870
## Year2001         -0.6931    0.3013  -2.30 0.02165 *
## Year2002         -0.8778    0.2761  -3.18 0.00152 **
## Year2003         -0.7281    0.2812  -2.59 0.00976 **
```

```

## Year2004          -0.5837      0.2947    -1.98   0.04788 *
## Year2005          -0.7972      0.2788    -2.86   0.00432 **
## Year2006          -0.8749      0.2861    -3.06   0.00228 **
## Year2007          -0.9434      0.2770    -3.41   0.00068 ***
## Year2008          -0.9008      0.2780    -3.24   0.00123 **
## Year2009          -0.6930      0.2802    -2.47   0.01356 *
## Year2010          -0.9560      0.2793    -3.42   0.00065 ***
## Year2011          -0.9438      0.2822    -3.34   0.00086 ***
## Year2012          -0.7205      0.2861    -2.52   0.01193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.696
## Multiple R-squared:  0.0761, Adjusted R-squared:  0.0608
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
##  57 weights are ~= 1. The remaining 992 ones are summarized as
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0113 0.8650 0.9380 0.8910 0.9740 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.53e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1049"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2912"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   11   13   15   11    9   20   12   11   15   25   33   23   28   27
## 2011 2012
##   33   30
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 9 7 13 7 4 17 10 8 14 20 29 18 23 21
## 2011 2012
## 29 26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 13 9 7 13 6 4 14 9 7 14 18 28 17 23 21
## 2011 2012
## 25 23
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```

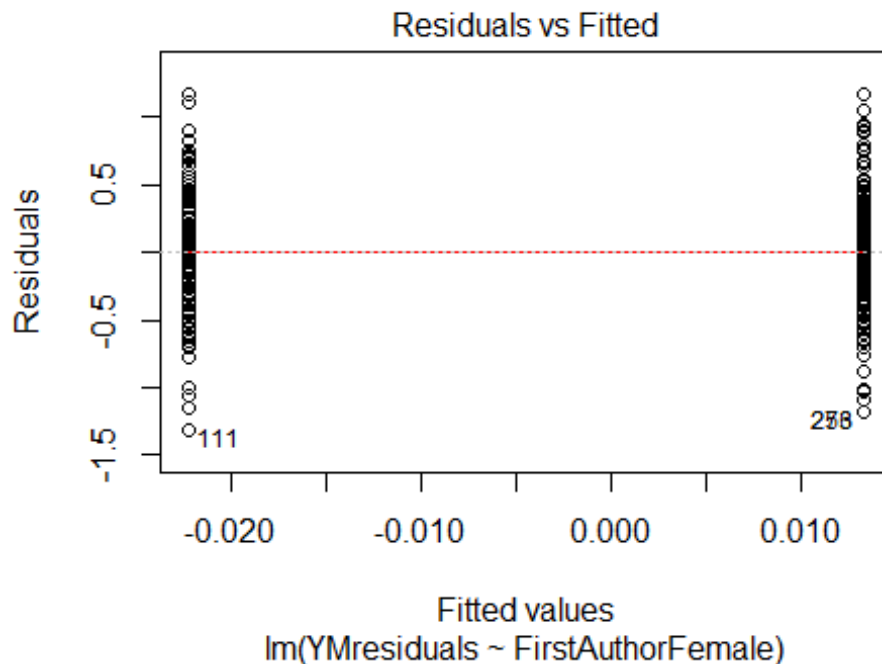


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.75, df = 1, p-value = 0.4
## [1] "Female first author team size 2018 geometric mean: 3.18528961423337"
## [1] "Male first author team size 2018 geometric mean: 2"
```

```
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 30, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.01760401644121"
## [1] "Male last author team size 2018 geometric mean: 1"

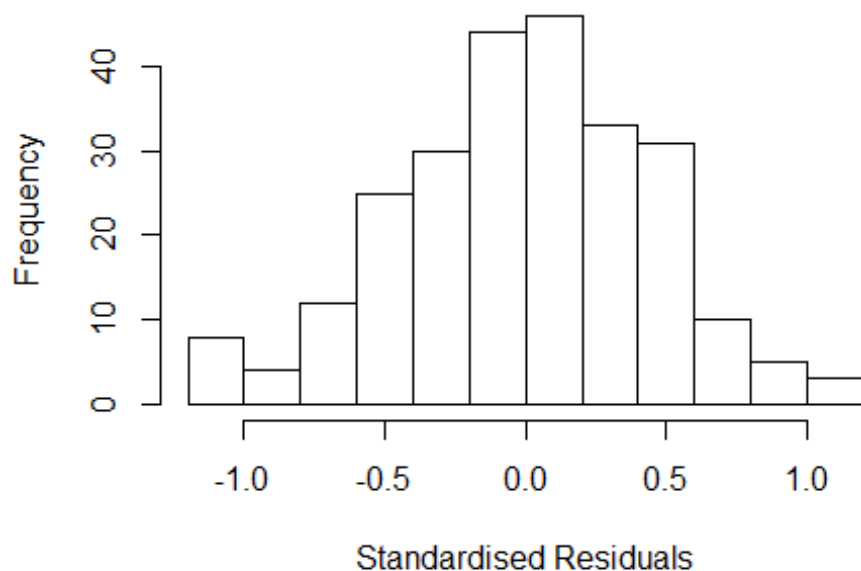
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
##      Year as factors"
##
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.416 1          1.190
```

## LastAuthorFemale	1.723	1	1.313
## UniqueAuthors	3.443	4	1.167
## Year	5.426	16	1.054

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1596 -0.2908 0.0114 0.3369 1.1913
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9709 0.2567 3.78 0.0002 ***
## FirstAuthorFemale1 -0.0494 0.0687 -0.72 0.4722
## LastAuthorFemale1 0.1808 0.0686 2.64 0.0090 **
## UniqueAuthors2 0.0563 0.0948 0.59 0.5531
## UniqueAuthors3 0.1386 0.1014 1.37 0.1728
## UniqueAuthors4 0.1698 0.1418 1.20 0.2322
```

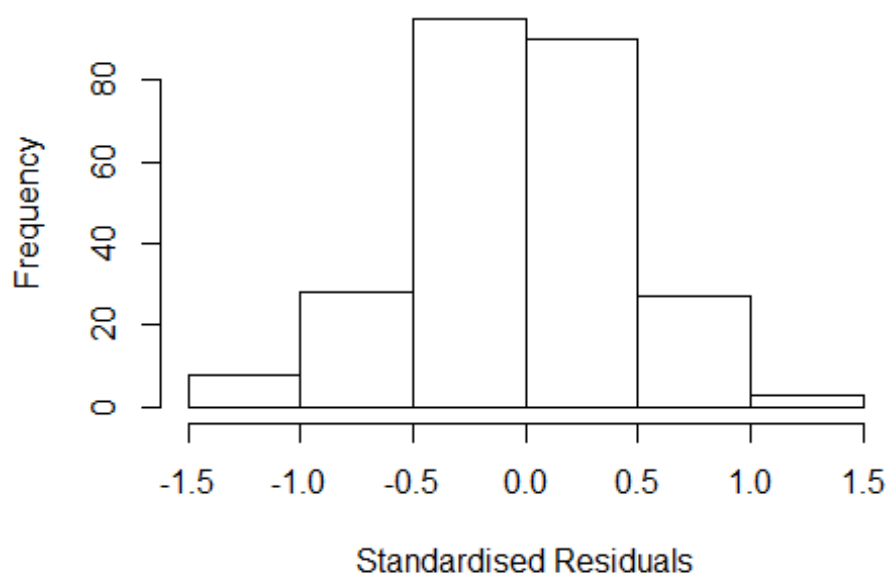
```

## UniqueAuthors5      0.2372      0.1200      1.98      0.0493 *
## Year1997             -0.6004      0.3054     -1.97      0.0505 .
## Year1998             -0.3845      0.3114     -1.23      0.2182
## Year1999              0.2935      0.3135      0.94      0.3501
## Year2000             -0.1150      0.3296     -0.35      0.7275
## Year2001             -0.3006      0.2931     -1.03      0.3061
## Year2002              0.1836      0.2825      0.65      0.5163
## Year2003              0.1294      0.3006      0.43      0.6673
## Year2004              0.1710      0.2946      0.58      0.5621
## Year2005              0.0499      0.2678      0.19      0.8522
## Year2006              0.0386      0.2711      0.14      0.8868
## Year2007              0.2841      0.2633      1.08      0.2819
## Year2008              0.0671      0.2717      0.25      0.8051
## Year2009              0.0573      0.2825      0.20      0.8394
## Year2010             -0.0983      0.2683     -0.37      0.7144
## Year2011             -0.0701      0.2720     -0.26      0.7968
## Year2012             -0.1043      0.2763     -0.38      0.7060
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.467
## Multiple R-squared:  0.191, Adjusted R-squared:  0.113
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 231 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.496  0.879  0.950  0.912  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.98e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.301 1      1.141
## LastAuthorFemale  1.509 1      1.228
## Year              1.799 16      1.019

```



## Residuals from first and last author



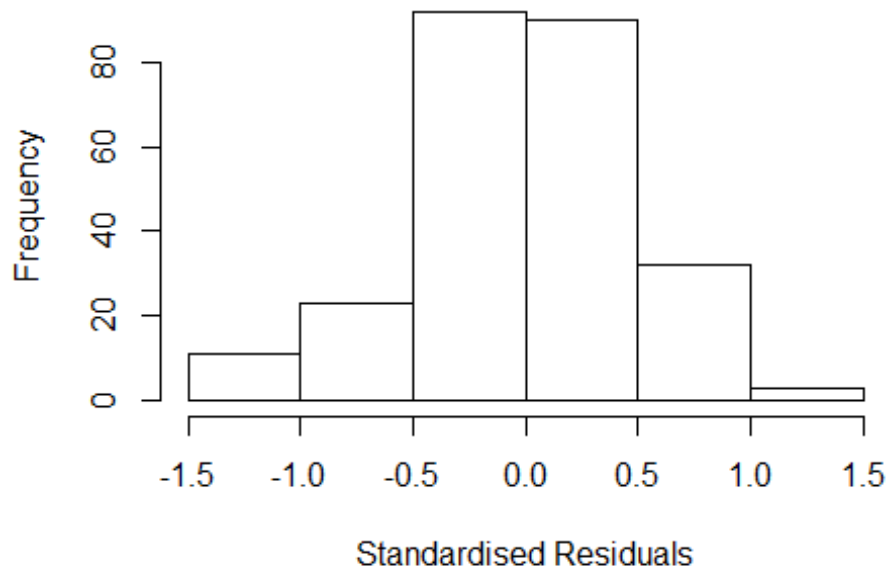
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3054 -0.3085 -0.0163 0.3151 1.2853
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0132 0.2709 3.74 0.00023 ***
## FirstAuthorFemale1 -0.0334 0.0669 -0.50 0.61799
## LastAuthorFemale1 0.1923 0.0677 2.84 0.00493 **
## Year1997 -0.6140 0.3206 -1.92 0.05670 .
## Year1998 -0.4227 0.3202 -1.32 0.18811
## Year1999 0.3180 0.3142 1.01 0.31256
## Year2000 -0.0734 0.3315 -0.22 0.82501
## Year2001 -0.3023 0.3034 -1.00 0.32007
## Year2002 0.2352 0.2882 0.82 0.41520
## Year2003 0.2202 0.3006 0.73 0.46464
## Year2004 0.2477 0.2986 0.83 0.40766
## Year2005 0.0619 0.2772 0.22 0.82360
```

```

## Year2006          0.0609      0.2776      0.22  0.82649
## Year2007          0.3205      0.2721      1.18  0.24012
## Year2008          0.1058      0.2782      0.38  0.70407
## Year2009          0.1333      0.2842      0.47  0.63946
## Year2010         -0.0847      0.2804     -0.30  0.76297
## Year2011         -0.0408      0.2767     -0.15  0.88297
## Year2012         -0.0343      0.2873     -0.12  0.90511
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.471
## Multiple R-squared:  0.179, Adjusted R-squared:  0.115
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 224 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.422  0.880   0.948   0.908   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.98e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.305 1      1.142
## Year              1.305 16      1.008

```

## Residuals from first author



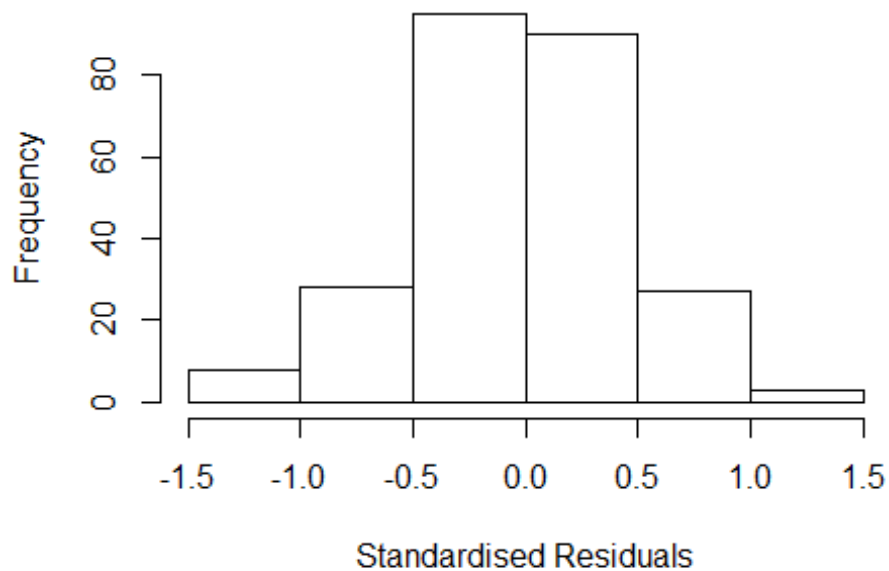
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3497 -0.3155 -0.0239  0.3415  1.2083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.03965    0.26084   3.99   9e-05 ***
## FirstAuthorFemale1 0.03098    0.06891   0.45   0.653
## Year1997      -0.51539    0.30835  -1.67   0.096 .
## Year1998      -0.31579    0.31285  -1.01   0.314
## Year1999       0.29534    0.29821   0.99   0.323
## Year2000      -0.02996    0.33913  -0.09   0.930
## Year2001      -0.23108    0.32028  -0.72   0.471
## Year2002       0.28854    0.28026   1.03   0.304
## Year2003       0.31000    0.28457   1.09   0.277
## Year2004       0.26792    0.28428   0.94   0.347
## Year2005       0.11457    0.26926   0.43   0.671
## Year2006       0.09423    0.27065   0.35   0.728
```

```

## Year2007          0.32051    0.26475    1.21    0.227
## Year2008          0.15958    0.27107    0.59    0.557
## Year2009          0.14722    0.27502    0.54    0.593
## Year2010         -0.04731    0.27456   -0.17    0.863
## Year2011         -0.02295    0.27105   -0.08    0.933
## Year2012         -0.00273    0.27421   -0.01    0.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.491
## Multiple R-squared:  0.144, Adjusted R-squared:  0.0814
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 232 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.430  0.891  0.952  0.916  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.98e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.448 1          1.204
## Year              1.448 16          1.012

```

## Residuals from last author



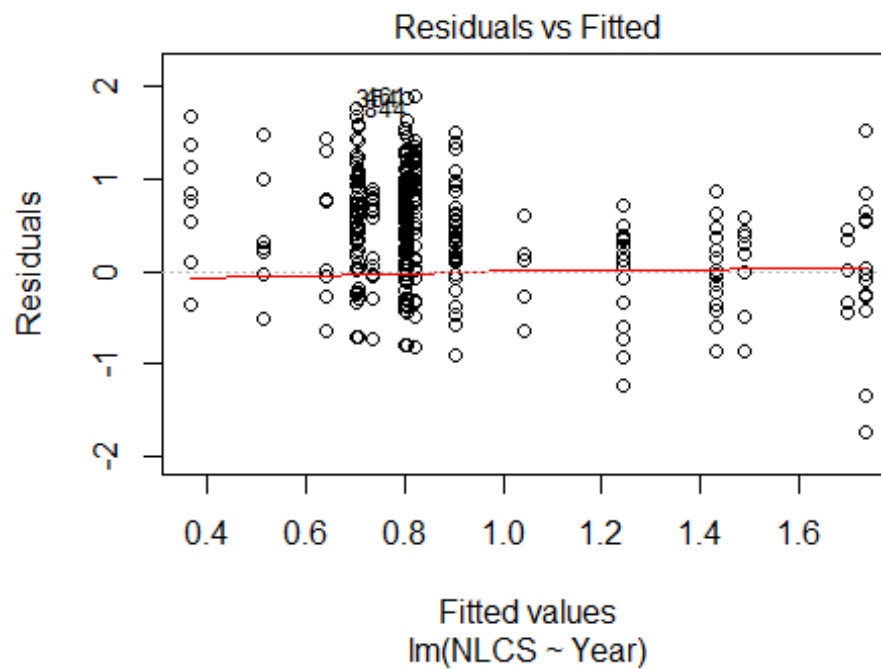
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3168 -0.3084 -0.0118 0.3147 1.2614
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9898 0.2649 3.74 0.00024 ***
## LastAuthorFemale1 0.1821 0.0673 2.71 0.00733 **
## Year1997 -0.6002 0.3180 -1.89 0.06033 .
## Year1998 -0.4193 0.3244 -1.29 0.19744
## Year1999 0.3183 0.3158 1.01 0.31453
## Year2000 -0.0581 0.3359 -0.17 0.86282
## Year2001 -0.2980 0.3100 -0.96 0.33752
## Year2002 0.2440 0.2882 0.85 0.39806
## Year2003 0.2241 0.3013 0.74 0.45774
## Year2004 0.2490 0.3001 0.83 0.40749
## Year2005 0.0709 0.2770 0.26 0.79812
## Year2006 0.0681 0.2784 0.24 0.80706
```

```

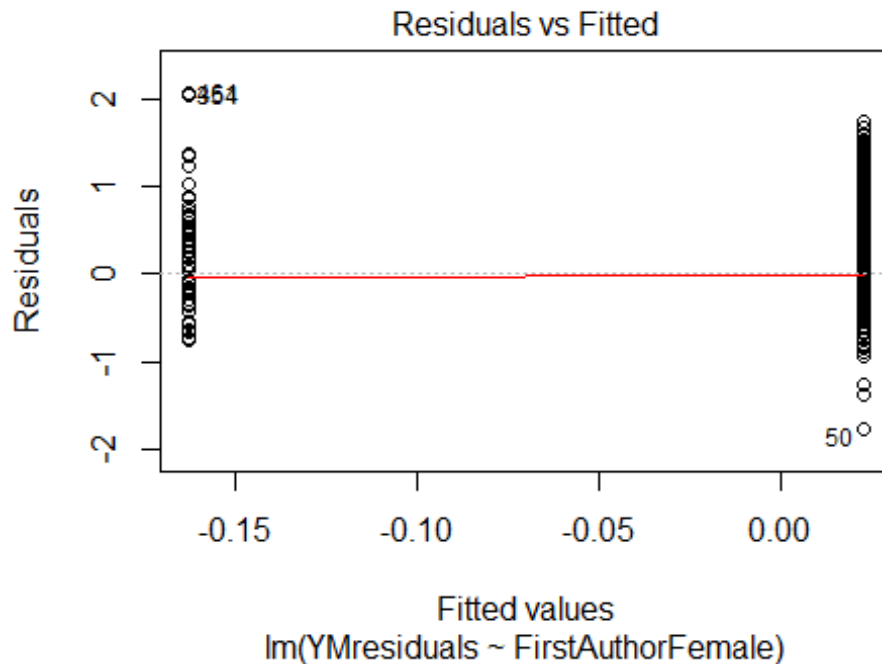
## Year2007          0.3289      0.2723      1.21  0.22834
## Year2008          0.1157      0.2790      0.41  0.67868
## Year2009          0.1448      0.2838      0.51  0.61033
## Year2010         -0.0749      0.2811     -0.27  0.79010
## Year2011         -0.0388      0.2786     -0.14  0.88939
## Year2012         -0.0222      0.2873     -0.08  0.93853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.179, Adjusted R-squared:  0.119
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 220 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.405  0.871  0.946  0.905  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.98e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 251"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2913"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   11   25   32   53   68   21   31   37   24  117  136   90  120  145
## 2011 2012
##  150  125
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    5   10   13   17   17   19   28   31   16  105  124   74  107  122
## 2011 2012

```

```
## 135 111
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 5 9 13 17 16 18 26 30 14 100 110 72 102 117
## 2011 2012
## 129 103
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.2
```



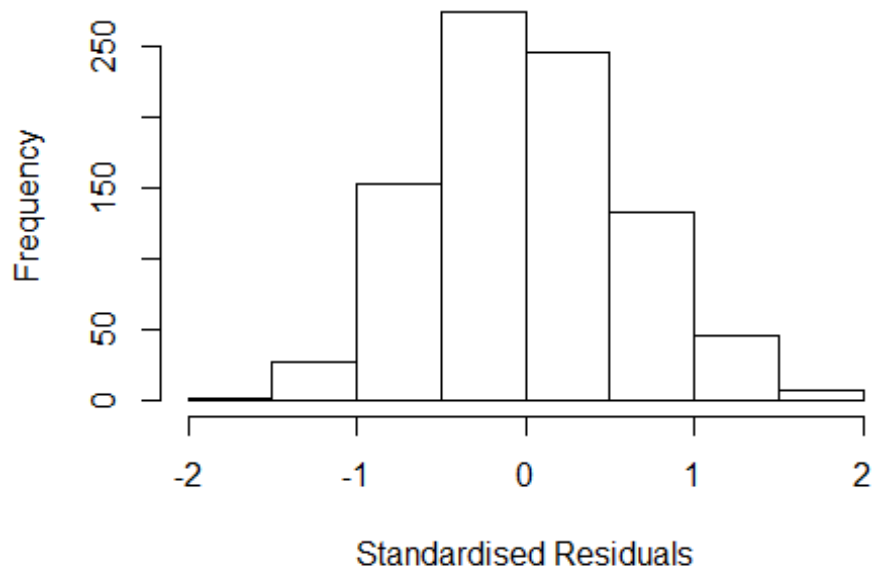
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.2, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 2.26839208454786"
## [1] "Male first author team size 2018 geometric mean: 2.02634602718282"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 450, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.17314345358576"
## [1] "Male last author team size 2018 geometric mean: 2.56439777609864"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 570, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.562 1      1.601
## LastAuthorFemale  2.626 1      1.620
## UniqueAuthors    1.634 4      1.063
## Year             1.897 16      1.020
```



## Residuals from first and last author and team size



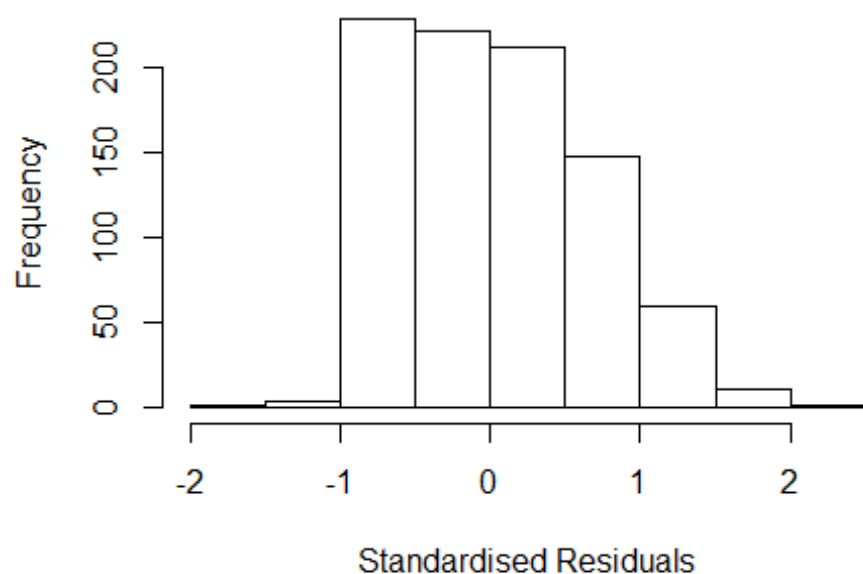
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5462 -0.4713 -0.0184 0.4297 1.8639
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6205 0.1886 3.29 0.0010 **
## FirstAuthorFemale1 0.1317 0.0863 1.53 0.1272
## LastAuthorFemale1 -0.0194 0.0813 -0.24 0.8114
## UniqueAuthors2 0.3781 0.0619 6.11 1.5e-09 ***
## UniqueAuthors3 0.5645 0.0720 7.84 1.4e-14 ***
## UniqueAuthors4 0.5032 0.1072 4.69 3.1e-06 ***
## UniqueAuthors5 0.6720 0.1153 5.83 7.9e-09 ***
## Year1997 0.7362 0.3319 2.22 0.0268 *
## Year1998 0.5258 0.2819 1.86 0.0625 .
## Year1999 0.8134 0.3303 2.46 0.0140 *
```

```

## Year2000          0.5173      0.2032      2.55      0.0111 *
## Year2001          0.2686      0.2192      1.23      0.2209
## Year2002         -0.3767      0.2069     -1.82      0.0691 .
## Year2003         -0.1878      0.2048     -0.92      0.3595
## Year2004         -0.5156      0.1976     -2.61      0.0092 **
## Year2005         -0.3977      0.2576     -1.54      0.1229
## Year2006         -0.1613      0.1942     -0.83      0.4063
## Year2007         -0.1261      0.1989     -0.63      0.5263
## Year2008         -0.0247      0.1985     -0.12      0.9011
## Year2009         -0.1058      0.1897     -0.56      0.5772
## Year2010         -0.2575      0.1913     -1.35      0.1785
## Year2011         -0.1639      0.1894     -0.87      0.3871
## Year2012         -0.2848      0.1934     -1.47      0.1413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.248, Adjusted R-squared:  0.228
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 70 weights are ~= 1. The remaining 816 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.313  0.887  0.944  0.910  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.13e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.519 1          1.587
## LastAuthorFemale 2.588 1          1.609
## Year          1.273 16          1.008

```

## Residuals from first and last author



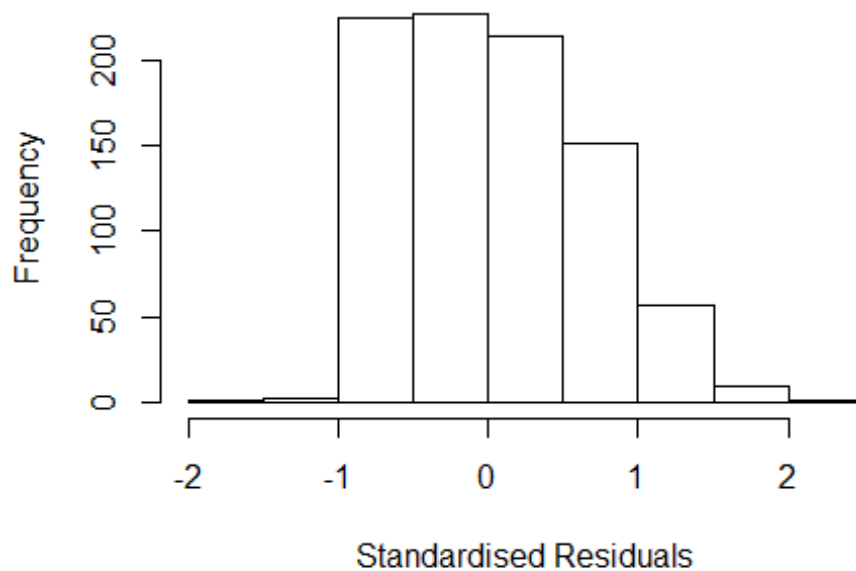
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7887 -0.5017 -0.0257 0.4853 2.2857
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8484 0.2791 3.04 0.00244 **
## FirstAuthorFemale1 0.3602 0.0949 3.80 0.00016 ***
## LastAuthorFemale1 -0.2012 0.0854 -2.36 0.01872 *
## Year1997 0.7598 0.3464 2.19 0.02857 *
## Year1998 0.4983 0.3397 1.47 0.14282
## Year1999 0.7812 0.3687 2.12 0.03440 *
## Year2000 0.4059 0.2853 1.42 0.15517
## Year2001 0.2415 0.3028 0.80 0.42533
## Year2002 -0.5063 0.2935 -1.72 0.08493 .
## Year2003 -0.2645 0.3014 -0.88 0.38037
## Year2004 -0.7424 0.2850 -2.60 0.00936 **
## Year2005 -0.5285 0.3345 -1.58 0.11447
```

```

## Year2006          -0.2628      0.2788   -0.94  0.34606
## Year2007          -0.2159      0.2814   -0.77  0.44314
## Year2008          -0.1033      0.2817   -0.37  0.71387
## Year2009          -0.1966      0.2763   -0.71  0.47690
## Year2010          -0.3475      0.2773   -1.25  0.21054
## Year2011          -0.2147      0.2772   -0.77  0.43877
## Year2012          -0.3233      0.2798   -1.16  0.24820
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.67
## Multiple R-squared:  0.13,   Adjusted R-squared:  0.112
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 79 weights are ~= 1. The remaining 807 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.221  0.879  0.940  0.916  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.093 1      1.045
## Year              1.093 16      1.003

```

## Residuals from first author



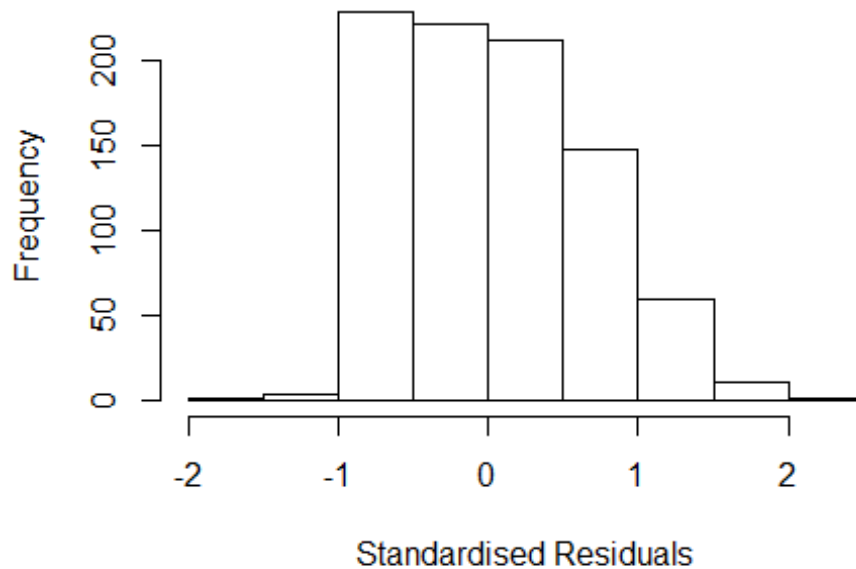
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8139 -0.5163 -0.0209 0.4889 2.0999
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8667 0.2403 3.61 0.00033 ***
## FirstAuthorFemale1 0.2134 0.0608 3.51 0.00047 ***
## Year1997 0.6606 0.2998 2.20 0.02785 *
## Year1998 0.4352 0.3075 1.42 0.15738
## Year1999 0.7338 0.3369 2.18 0.02969 *
## Year2000 0.3778 0.2543 1.49 0.13776
## Year2001 0.2006 0.2732 0.73 0.46298
## Year2002 -0.5688 0.2568 -2.21 0.02703 *
## Year2003 -0.3309 0.2663 -1.24 0.21428
## Year2004 -0.8064 0.2471 -3.26 0.00114 **
## Year2005 -0.5622 0.3082 -1.82 0.06849 .
## Year2006 -0.3097 0.2421 -1.28 0.20113
```

```

## Year2007          -0.2496      0.2457   -1.02   0.30986
## Year2008          -0.1623      0.2434   -0.67   0.50509
## Year2009          -0.2454      0.2389   -1.03   0.30476
## Year2010          -0.4021      0.2390   -1.68   0.09276 .
## Year2011          -0.2742      0.2382   -1.15   0.25009
## Year2012          -0.3940      0.2404   -1.64   0.10162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.675
## Multiple R-squared:  0.123, Adjusted R-squared:  0.106
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 811 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.313  0.877  0.943  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.134 1          1.065
## Year              1.134 16          1.004

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.7929 -0.6046 -0.0248 0.4741 1.9430
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04756 0.20144 5.20 2.5e-07 ***
## LastAuthorFemale1 -0.00442 0.06296 -0.07 0.94403
## Year1997 0.65590 0.26222 2.50 0.01256 *
## Year1998 0.44559 0.28027 1.59 0.11223
## Year1999 0.74977 0.31008 2.42 0.01581 *
## Year2000 0.38598 0.22536 1.71 0.08711 .
## Year2001 0.21260 0.24980 0.85 0.39496
## Year2002 -0.57006 0.23091 -2.47 0.01375 *
## Year2003 -0.30849 0.24203 -1.27 0.20280
## Year2004 -0.76967 0.22167 -3.47 0.00054 ***
## Year2005 -0.56719 0.28077 -2.02 0.04368 *
## Year2006 -0.29936 0.21395 -1.40 0.16212
```

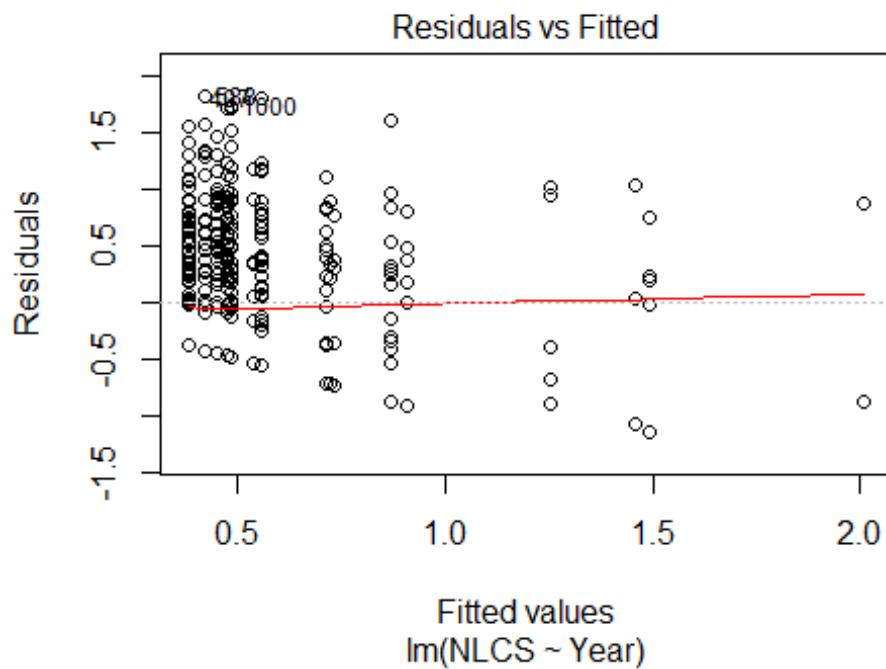
```

## Year2007          -0.26914      0.21757    -1.24   0.21641
## Year2008          -0.14756      0.21554    -0.68   0.49378
## Year2009          -0.23176      0.21062    -1.10   0.27148
## Year2010          -0.39860      0.21085    -1.89   0.05903 .
## Year2011          -0.26042      0.21000    -1.24   0.21528
## Year2012          -0.37905      0.21268    -1.78   0.07505 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.675
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0962
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 54 weights are ~= 1. The remaining 832 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.387  0.882  0.941  0.918  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.13e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 886"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2914"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12   6   8  11  12  11  14  22  18  25 113  80  83 122 107
## 2011 2012
##  129  118
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   3   2   5   5   7   5  11  19   6  18  97  64  71 102  95
## 2011 2012

```



```
## 116 99
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 1 4 5 7 5 11 16 6 17 92 62 69 95 92
## 2011 2012
## 110 96
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 23, df = 16, p-value = 0.1
```



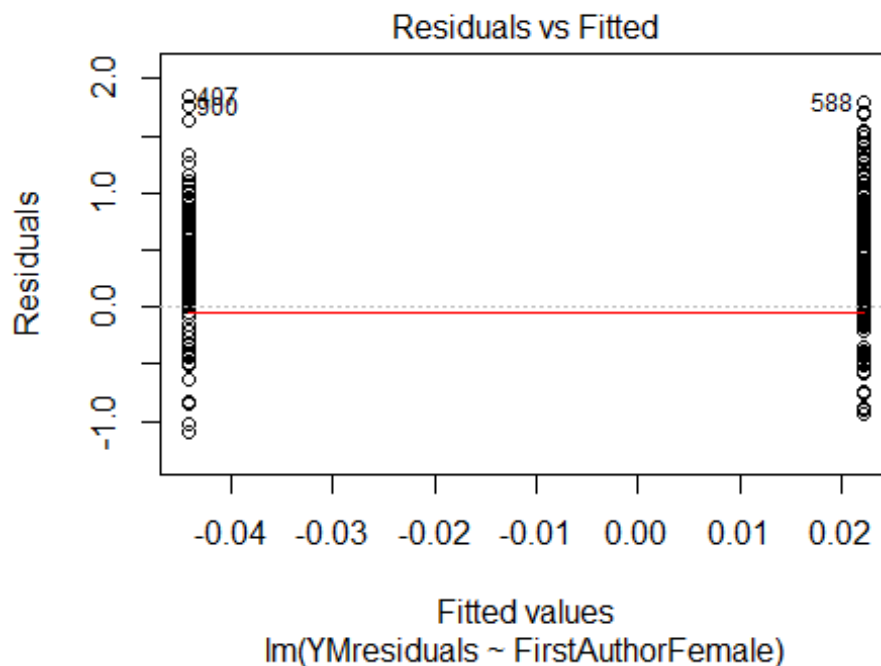
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 2.55427729879623"
## [1] "Male first author team size 2018 geometric mean: 2.88226049223315"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 440, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.47508954163807"
## [1] "Male last author team size 2018 geometric mean: 3.31466054471101"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale  6.43  1      2.536
## LastAuthorFemale 11.34  1      3.368
## UniqueAuthors    80.10  4      1.730
## Year             100.54 16      1.155
```

```

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.8327 -0.2905 -0.0905 0.3586 2.0146
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.9199 0.3012 6.37 3.4e-10 ***
## FirstAuthorFemale1 0.1516 0.0727 2.09 0.03726 *
## LastAuthorFemale1 -0.1465 0.0765 -1.91 0.05599 .
## UniqueAuthors2 0.0676 0.0557 1.21 0.22546
## UniqueAuthors3 0.3321 0.1375 2.42 0.01597 *
## UniqueAuthors4 0.5297 0.1223 4.33 1.7e-05 ***
## UniqueAuthors5 0.6435 0.1165 5.52 4.8e-08 ***
## Year1997 -1.5810 0.3197 -4.94 9.7e-07 ***
## Year1998 0.2798 0.3104 0.90 0.36765
## Year1999 -0.4323 0.4967 -0.87 0.38442
## Year2000 -0.9906 0.4943 -2.00 0.04546 *
## Year2001 -1.4128 0.5438 -2.60 0.00958 **
## Year2002 -1.5461 0.3479 -4.44 1.0e-05 ***
## Year2003 -1.4751 0.3291 -4.48 8.7e-06 ***
## Year2004 -1.6202 0.3721 -4.35 1.5e-05 ***
## Year2005 -1.2883 0.3794 -3.40 0.00073 ***
## Year2006 -1.6910 0.3030 -5.58 3.5e-08 ***
## Year2007 -1.7028 0.3063 -5.56 3.9e-08 ***
## Year2008 -1.5765 0.3084 -5.11 4.2e-07 ***
## Year2009 -1.6209 0.3052 -5.31 1.5e-07 ***
## Year2010 -1.6645 0.3042 -5.47 6.3e-08 ***
## Year2011 -1.6294 0.3052 -5.34 1.3e-07 ***
## Year2012 -1.6558 0.3064 -5.40 9.1e-08 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.382
## Multiple R-squared: 0.292, Adjusted R-squared: 0.269
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 3 observations c(7,262,397) are outliers with |weight| = 0 ( < 0.00014);
## 15 weights are ~ = 1. The remaining 672 ones are summarized as

```

```

##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## 0.0024 0.8480 0.9450 0.8690 0.9680 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.45e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate

## [1] "Regression 3: First author gender, Year as factors"

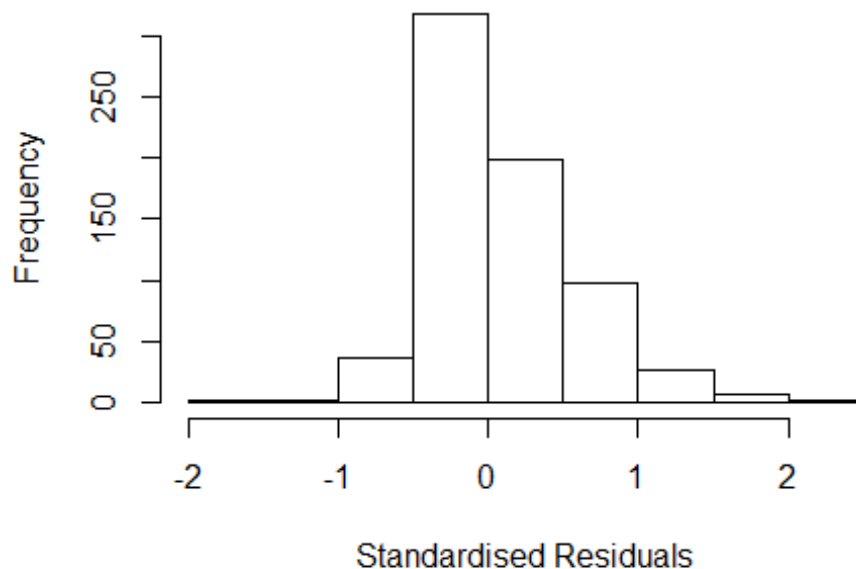
## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate

## [1] "Regression 4: Last author gender, Year as factors"

## Warning in lmrob.fit(x, y, control, init = init, mf = mf): M-step did NOT
## converge. Returning unconverged SM-estimate

```

## Residuals from first and last author and team size

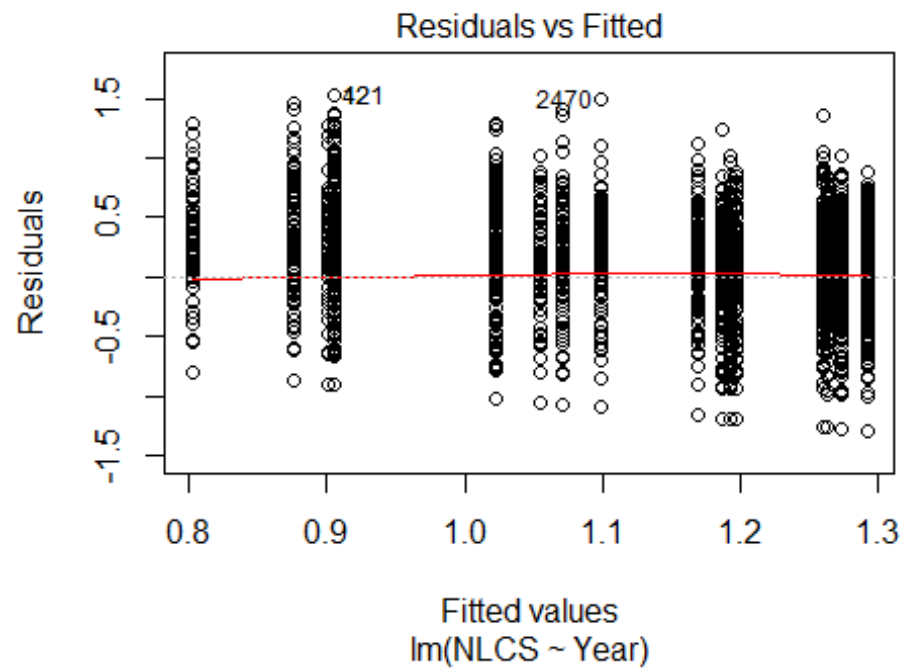


```
## [1] "Sample size for the above analysis: 690"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2915"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2006 2009 2011
##    1    9    1
##
## 2006 2009 2011
##    1    8    0
##
## 2006 2009 2011
##    1    8    0
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.5874010519682"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
```

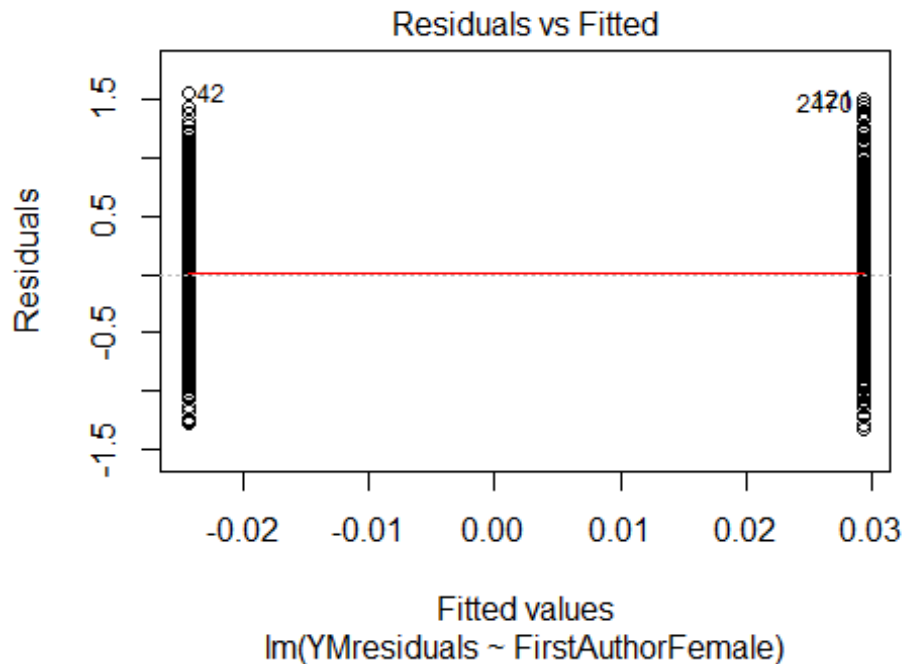
```

## [1] "Sample size for the above analysis: 9"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2916"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 293 275 309 238 291 251 219 161 232 188 192 227 253 326 295
## 2011 2012
## 367 347
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 172 147 169 122 126 79 106 85 132 134 116 140 166 196 178
## 2011 2012
## 212 188
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 157 139 154 101 121 73 92 70 102 102 98 123 149 171 165
## 2011 2012
## 189 179
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16

```



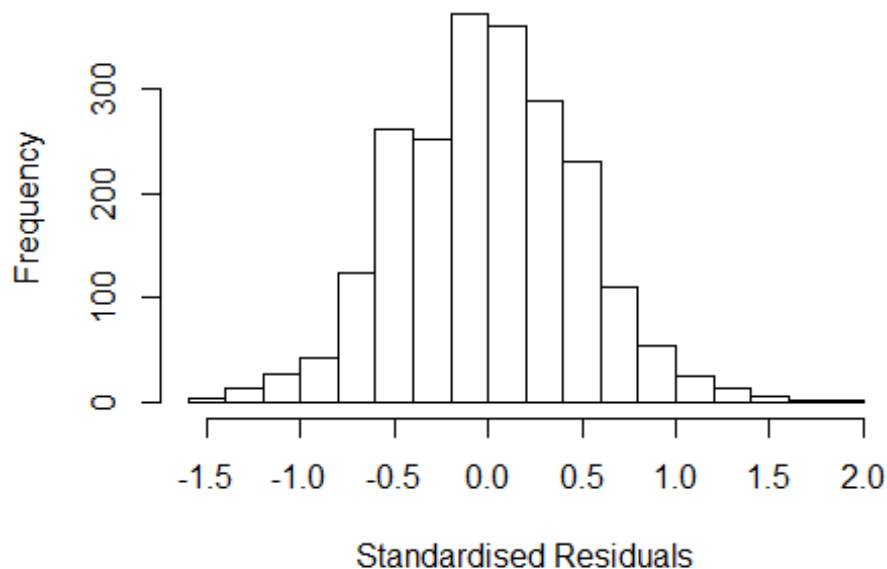
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 4.2907633640503"
## [1] "Male first author team size 2018 geometric mean: 4.65633238166774"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.37420385719126"
## [1] "Male last author team size 2018 geometric mean: 4.44061001820783"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5300, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.086 1      1.042
## LastAuthorFemale  1.118 1      1.058
## UniqueAuthors    1.390 4      1.042
## Year              1.447 16     1.012
```



## Residuals from first and last author and team size



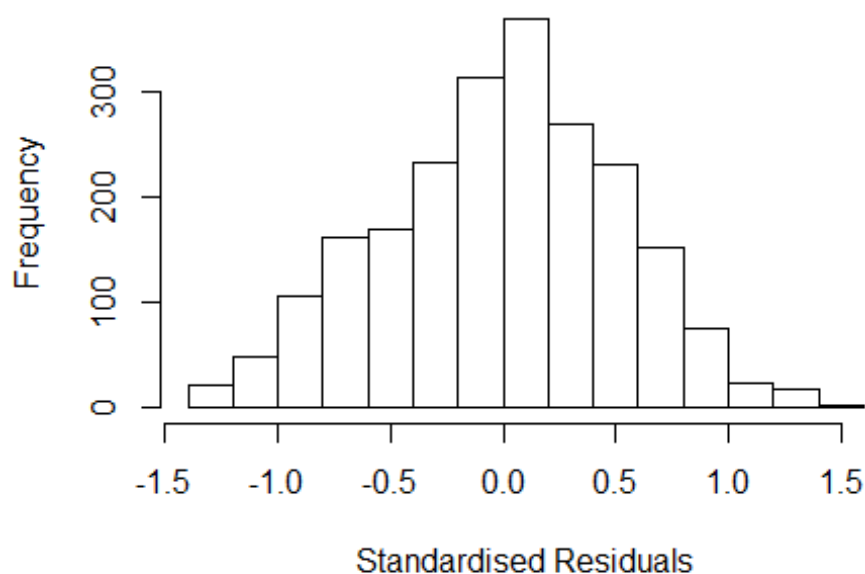
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.43081 -0.33193 -0.00192  0.32105  1.88208
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.59196    0.05872   10.08  <2e-16 ***
## FirstAuthorFemale1 -0.02570    0.02154   -1.19  0.2330
## LastAuthorFemale1 -0.07064    0.02269   -3.11  0.0019 **
## UniqueAuthors2     0.46712    0.04194   11.14  <2e-16 ***
## UniqueAuthors3     0.52298    0.03997   13.08  <2e-16 ***
## UniqueAuthors4     0.57037    0.04283   13.32  <2e-16 ***
## UniqueAuthors5     0.69589    0.03723   18.69  <2e-16 ***
## Year1997          -0.00589    0.07478   -0.08  0.9372
## Year1998           0.11119    0.07242    1.54  0.1248
## Year1999          -0.05358    0.07224   -0.74  0.4584
```

```

## Year2000      -0.09164      0.06779      -1.35      0.1766
## Year2001      -0.02726      0.07960      -0.34      0.7320
## Year2002       0.10794      0.07695       1.40      0.1609
## Year2003       0.04118      0.07383       0.56      0.5771
## Year2004       0.08280      0.07292       1.14      0.2563
## Year2005       0.12030      0.06597       1.82      0.0684 .
## Year2006       0.17680      0.07231       2.44      0.0146 *
## Year2007       0.12992      0.06699       1.94      0.0526 .
## Year2008       0.11675      0.06712       1.74      0.0821 .
## Year2009       0.17930      0.06446       2.78      0.0055 **
## Year2010       0.20062      0.06494       3.09      0.0020 **
## Year2011       0.19561      0.06373       3.07      0.0022 **
## Year2012       0.14296      0.06435       2.22      0.0264 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.457
## Multiple R-squared:  0.298, Adjusted R-squared:  0.291
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 175 weights are ~= 1. The remaining 2010 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0512 0.8660 0.9430 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.58e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.056 1      1.028
## LastAuthorFemale  1.076 1      1.037
## Year              1.089 16      1.003

```

## Residuals from first and last author



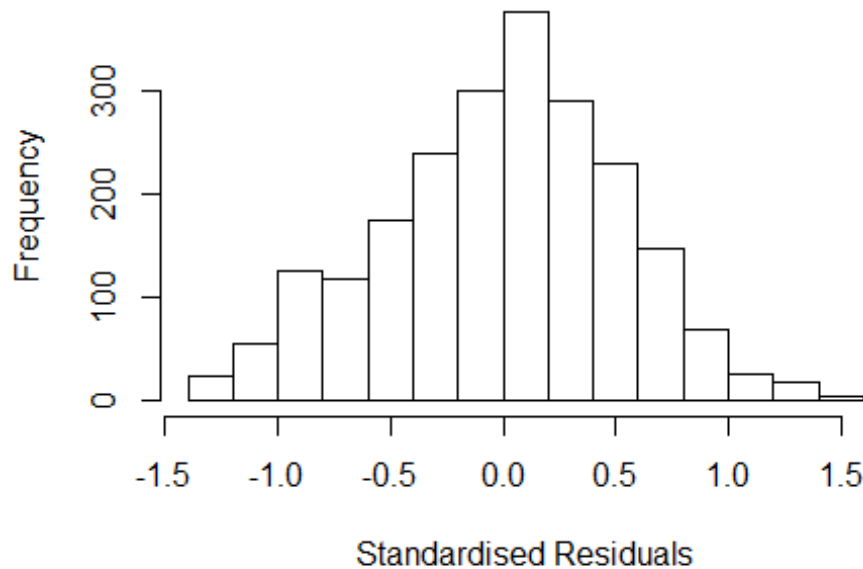
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3650 -0.3579  0.0253  0.3653  1.5163
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91366    0.06112   14.95 < 2e-16 ***
## FirstAuthorFemale1 -0.00627    0.02379   -0.26  0.79227
## LastAuthorFemale1 -0.14740    0.02516   -5.86  5.3e-09 ***
## Year1997        -0.01616    0.08397   -0.19  0.84742
## Year1998         0.16934    0.08158    2.08  0.03803 *
## Year1999        -0.03152    0.08666   -0.36  0.71606
## Year2000        -0.07504    0.08636   -0.87  0.38494
## Year2001         0.01577    0.10239    0.15  0.87762
## Year2002         0.24661    0.08916    2.77  0.00573 **
## Year2003         0.23505    0.08822    2.66  0.00777 **
## Year2004         0.32252    0.08604    3.75  0.00018 ***
## Year2005         0.33232    0.07429    4.47  8.1e-06 ***
```

```

## Year2006          0.40494    0.07620    5.31  1.2e-07 ***
## Year2007          0.33218    0.07549    4.40  1.1e-05 ***
## Year2008          0.36181    0.07504    4.82  1.5e-06 ***
## Year2009          0.43448    0.06804    6.39  2.1e-10 ***
## Year2010          0.45135    0.06989    6.46  1.3e-10 ***
## Year2011          0.43806    0.06842    6.40  1.9e-10 ***
## Year2012          0.40867    0.06967    5.87  5.2e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.122, Adjusted R-squared:  0.115
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2000 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.371  0.858   0.946   0.905   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.043 1         1.021
## Year              1.043 16         1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3291 -0.3641  0.0265  0.3628  1.5606
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.86945    0.06312   13.77 < 2e-16 ***
## FirstAuthorFemale1 -0.04507    0.02401   -1.88  0.0607 .
## Year1997         0.00208    0.08742    0.02  0.9810
## Year1998         0.18425    0.08470    2.18  0.0297 *
## Year1999        -0.02828    0.09123   -0.31  0.7566
## Year2000        -0.07143    0.08980   -0.80  0.4264
## Year2001         0.01425    0.10490    0.14  0.8919
## Year2002         0.25982    0.09149    2.84  0.0046 **
## Year2003         0.24458    0.09064    2.70  0.0070 **
## Year2004         0.35630    0.08870    4.02 6.1e-05 ***
## Year2005         0.35637    0.07743    4.60 4.4e-06 ***
## Year2006         0.40136    0.07847    5.11 3.4e-07 ***
```

```

## Year2007          0.34748    0.07699    4.51  6.7e-06 ***
## Year2008          0.37073    0.07698    4.82  1.6e-06 ***
## Year2009          0.44084    0.07043    6.26  4.6e-10 ***
## Year2010          0.45966    0.07245    6.34  2.7e-10 ***
## Year2011          0.43587    0.07040    6.19  7.1e-10 ***
## Year2012          0.41357    0.07175    5.76  9.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.518
## Multiple R-squared:  0.108, Adjusted R-squared:  0.101
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 211 weights are ~= 1. The remaining 1974 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.344  0.855  0.947   0.902  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.057 1      1.028
## Year      1.057 16      1.002

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.3622 -0.3589  0.0239  0.3666  1.5182

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9118    0.0608   14.99 < 2e-16 ***
## LastAuthorFemale1 -0.1492    0.0251   -5.94  3.4e-09 ***
## Year1997          -0.0165    0.0840   -0.20  0.84460
## Year1998           0.1689    0.0817    2.07  0.03886 *
## Year1999          -0.0325    0.0865   -0.38  0.70684
## Year2000          -0.0760    0.0863   -0.88  0.37863
## Year2001           0.0147    0.1020    0.14  0.88570
## Year2002           0.2461    0.0892    2.76  0.00585 **
## Year2003           0.2340    0.0881    2.66  0.00797 **
## Year2004           0.3212    0.0859    3.74  0.00019 ***
## Year2005           0.3311    0.0741    4.47  8.4e-06 ***
## Year2006           0.4039    0.0761    5.30  1.2e-07 ***
## Year2007           0.3313    0.0754    4.39  1.2e-05 ***
## Year2008           0.3606    0.0748    4.82  1.5e-06 ***
## Year2009           0.4336    0.0679    6.38  2.1e-10 ***
## Year2010           0.4504    0.0697    6.46  1.3e-10 ***
## Year2011           0.4370    0.0683    6.39  2.0e-10 ***
## Year2012           0.4080    0.0697    5.85  5.5e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.517
## Multiple R-squared:  0.122, Adjusted R-squared:  0.115
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 2002 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.369  0.857   0.947   0.905   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2185"
## [1] ""

```

```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2917"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1   33   28   26   19   16   19   11   12   14   40   49   42   60   43
## 2011 2012
##    77   61
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1   31   25   19   15   12   15    9   12   13   31   37   39   49   38
## 2011 2012
##    67   48
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    1   30   25   19   14   12   15    9   11   10   29   33   38   45   37
## 2011 2012
##    59   44
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.11806792942181"
## [1] "Male first author team size 2018 geometric mean: 2.70213350320354"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

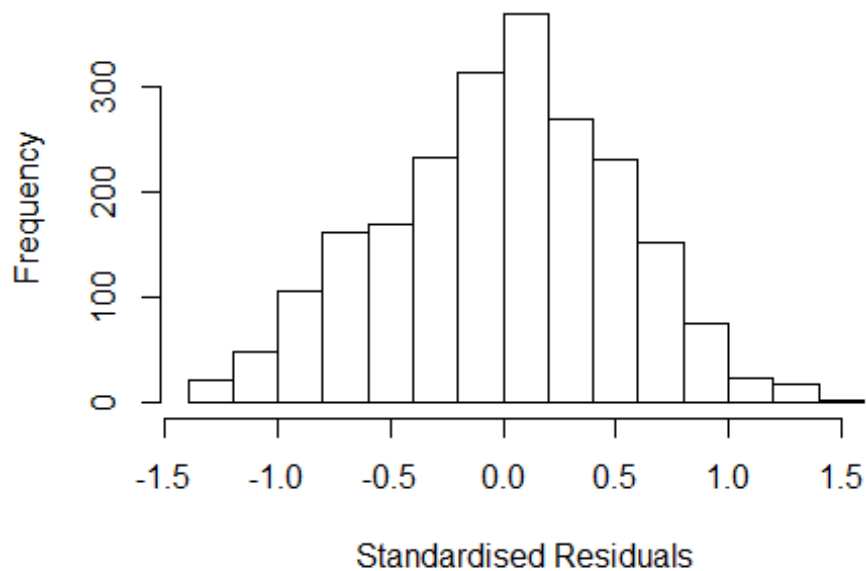
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.82876274914824"
## [1] "Male last author team size 2018 geometric mean: 3.21120783068661"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```

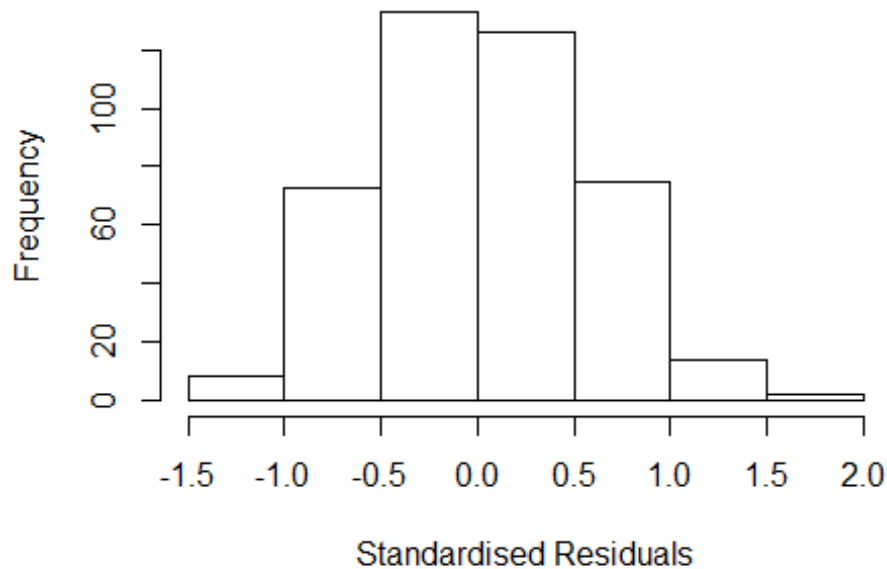


## Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 220, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.109 1      1.452
## LastAuthorFemale  2.187 1      1.479
## UniqueAuthors    8.303 4      1.303
## Year             11.844 16      1.080
```

## Residuals from first and last author and team size



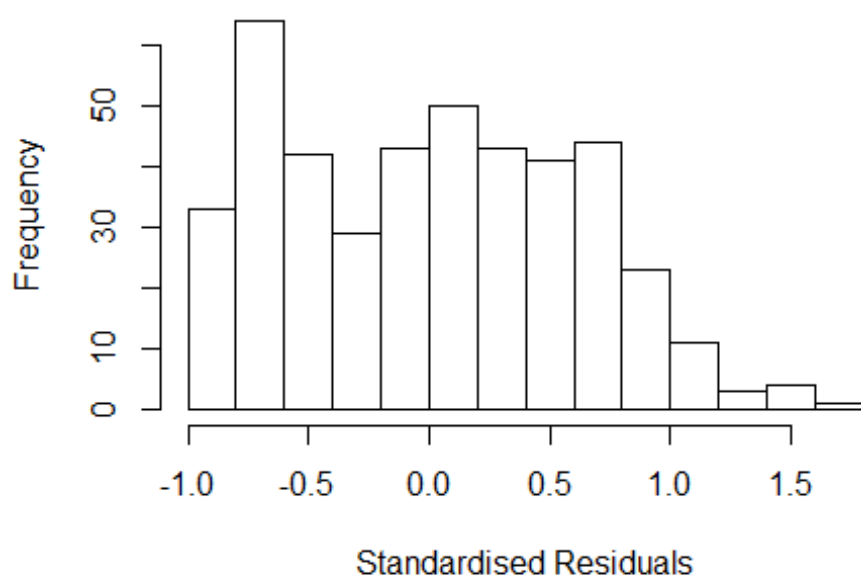
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28397 -0.43284 0.00706 0.40390 1.76826
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.15558 0.08500 13.59 < 2e-16 ***
## FirstAuthorFemale1 0.00353 0.09096 0.04 0.96907
## LastAuthorFemale1 0.06189 0.08103 0.76 0.44543
## UniqueAuthors2 0.14562 0.07762 1.88 0.06136 .
## UniqueAuthors3 0.52709 0.12256 4.30 2.1e-05 ***
## UniqueAuthors4 0.68172 0.09556 7.13 4.5e-12 ***
## UniqueAuthors5 0.83395 0.11656 7.15 3.9e-12 ***
## Year1997 -0.78816 0.12394 -6.36 5.4e-10 ***
## Year1998 -0.52563 0.08201 -6.41 4.0e-10 ***
## Year1999 -0.38205 0.13479 -2.83 0.00482 **
```

```

## Year2000      -0.46953    0.18033    -2.60    0.00956 **
## Year2001      -0.44411    0.12636    -3.51    0.00049 ***
## Year2002      -0.75557    0.11926    -6.34    6.3e-10 ***
## Year2003      -0.85094    0.17574    -4.84    1.8e-06 ***
## Year2004      -0.46145    0.19031    -2.42    0.01575 *
## Year2005      -0.35291    0.10428    -3.38    0.00078 ***
## Year2006      -0.66773    0.13122    -5.09    5.5e-07 ***
## Year2007      -0.79341    0.12930    -6.14    2.0e-09 ***
## Year2008      -0.69885    0.14682    -4.76    2.7e-06 ***
## Year2009      -0.53156    0.12357    -4.30    2.1e-05 ***
## Year2010      -0.75726    0.11057    -6.85    2.8e-11 ***
## Year2011      -0.77099    0.09591    -8.04    9.9e-15 ***
## Year2012      -0.90045    0.09825    -9.17    < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.225, Adjusted R-squared:  0.183
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 32 weights are ~= 1. The remaining 399 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.286  0.872  0.943   0.906  0.979   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.32e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.935 1 1.391
## LastAuthorFemale 2.216 1 1.489
## Year 2.069 16 1.023

```

## Residuals from first and last author



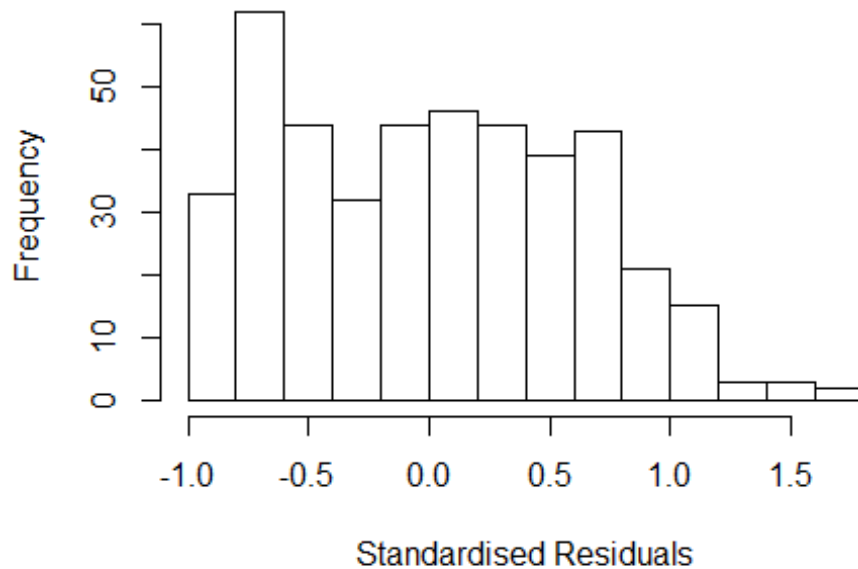
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.93529 -0.56174 0.00829 0.48743 1.68471
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0876 0.0888 12.25 < 2e-16 ***
## FirstAuthorFemale1 0.1008 0.0864 1.17 0.24383
## LastAuthorFemale1 0.0326 0.0823 0.40 0.69253
## Year1997 -0.6593 0.1268 -5.20 3.2e-07 ***
## Year1998 -0.5056 0.0882 -5.74 1.9e-08 ***
## Year1999 -0.3412 0.1322 -2.58 0.01018 *
## Year2000 -0.2976 0.1810 -1.64 0.10080
## Year2001 -0.3175 0.1348 -2.35 0.01902 *
## Year2002 -0.3942 0.1100 -3.58 0.00038 ***
## Year2003 -0.5847 0.1381 -4.23 2.8e-05 ***
## Year2004 -0.1275 0.1230 -1.04 0.30034
## Year2005 0.0686 0.1558 0.44 0.65976
```

```

## Year2006          -0.2857      0.1499    -1.91   0.05743 .
## Year2007          -0.4473      0.1493    -3.00   0.00290 **
## Year2008          -0.3305      0.1368    -2.42   0.01611 *
## Year2009          -0.3393      0.1206    -2.81   0.00513 **
## Year2010          -0.5117      0.1200    -4.26   2.5e-05 ***
## Year2011          -0.4502      0.0919    -4.90   1.4e-06 ***
## Year2012          -0.5394      0.1075    -5.02   7.8e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.0502, Adjusted R-squared:  0.00869
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 385 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.524  0.888   0.937   0.922   0.979   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.32e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.529 1      1.237
## Year      1.529 16      1.013

```

## Residuals from first author



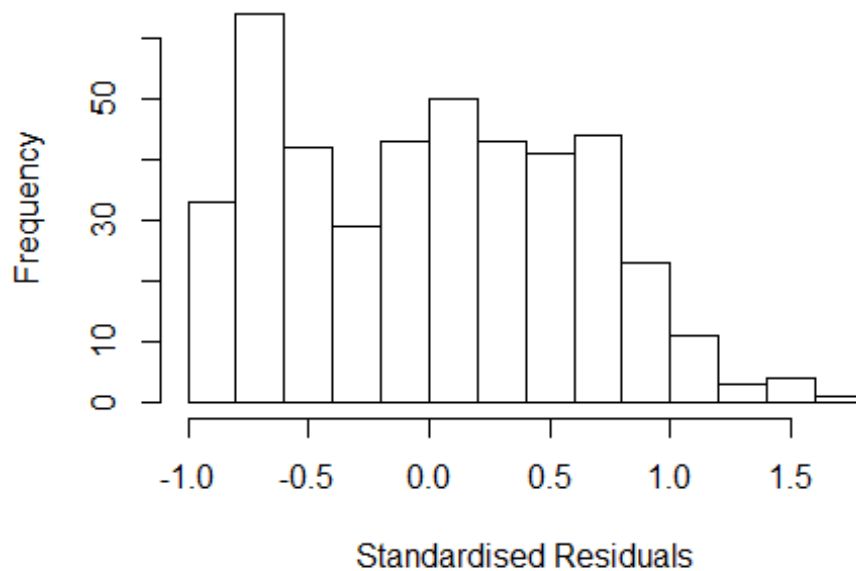
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.93357 -0.56230 0.00545 0.49474 1.68643
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1062 0.0771 14.35 < 2e-16 ***
## FirstAuthorFemale1 0.1148 0.0771 1.49 0.13733
## Year1997 -0.6587 0.1265 -5.21 3.0e-07 ***
## Year1998 -0.5077 0.0879 -5.78 1.5e-08 ***
## Year1999 -0.3454 0.1328 -2.60 0.00962 **
## Year2000 -0.2988 0.1795 -1.66 0.09670 .
## Year2001 -0.3225 0.1338 -2.41 0.01636 *
## Year2002 -0.3961 0.1098 -3.61 0.00035 ***
## Year2003 -0.5870 0.1375 -4.27 2.4e-05 ***
## Year2004 -0.1285 0.1226 -1.05 0.29506
## Year2005 0.0650 0.1537 0.42 0.67251
## Year2006 -0.2874 0.1486 -1.93 0.05374 .
```

```

## Year2007          -0.4541      0.1466   -3.10   0.00208 **
## Year2008          -0.3415      0.1293   -2.64   0.00856 **
## Year2009          -0.3475      0.1156   -3.01   0.00281 **
## Year2010          -0.5176      0.1178   -4.39   1.4e-05 ***
## Year2011          -0.4583      0.0891   -5.14   4.2e-07 ***
## Year2012          -0.5467      0.1044   -5.24   2.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.699
## Multiple R-squared:  0.0494, Adjusted R-squared:  0.0103
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 45 weights are ~= 1. The remaining 386 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.540  0.894  0.940  0.926  0.979  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.32e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.726 1      1.314
## Year              1.726 16      1.017

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.91641 -0.56211 0.00545 0.48811 1.70359
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1461 0.0739 15.51 < 2e-16 ***
## LastAuthorFemale1 0.0749 0.0739 1.01 0.31137
## Year1997 -0.6589 0.1248 -5.28 2.1e-07 ***
## Year1998 -0.5125 0.0877 -5.84 1.0e-08 ***
## Year1999 -0.3407 0.1296 -2.63 0.00888 **
## Year2000 -0.3154 0.1794 -1.76 0.07944 .
## Year2001 -0.3300 0.1271 -2.60 0.00978 **
## Year2002 -0.3989 0.1105 -3.61 0.00034 ***
## Year2003 -0.6084 0.1407 -4.32 1.9e-05 ***
## Year2004 -0.1427 0.1201 -1.19 0.23520
## Year2005 0.0564 0.1582 0.36 0.72173
## Year2006 -0.3046 0.1546 -1.97 0.04945 *
```

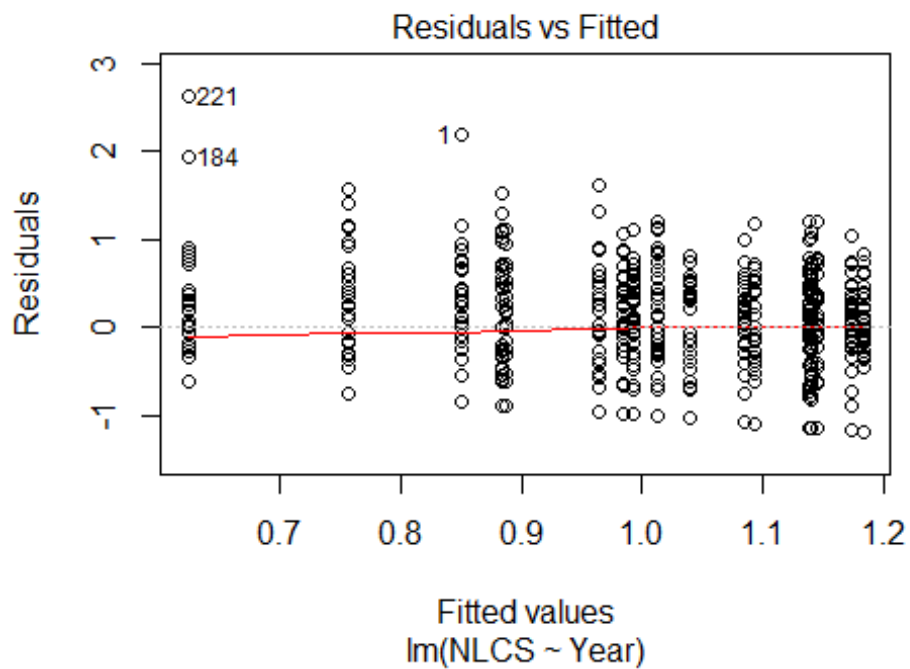


```

## Year2007          -0.4519      0.1447    -3.12   0.00192 **
## Year2008          -0.3401      0.1364    -2.49   0.01304 *
## Year2009          -0.3554      0.1200    -2.96   0.00323 **
## Year2010          -0.5365      0.1180    -4.55   7.2e-06 ***
## Year2011          -0.4657      0.0920    -5.06   6.3e-07 ***
## Year2012          -0.5539      0.1052    -5.26   2.3e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.696
## Multiple R-squared:  0.0466, Adjusted R-squared:  0.00737
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 390 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.529  0.892  0.935   0.926  0.978   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.32e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 431"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2919"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   56   61   33   62   45   37   46   48   33   38   37   47   45   51   63
## 2011 2012
##   51   39
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   49   49   27   57   40   34   43   41   28   32   34   42   41   44   56
## 2011 2012

```

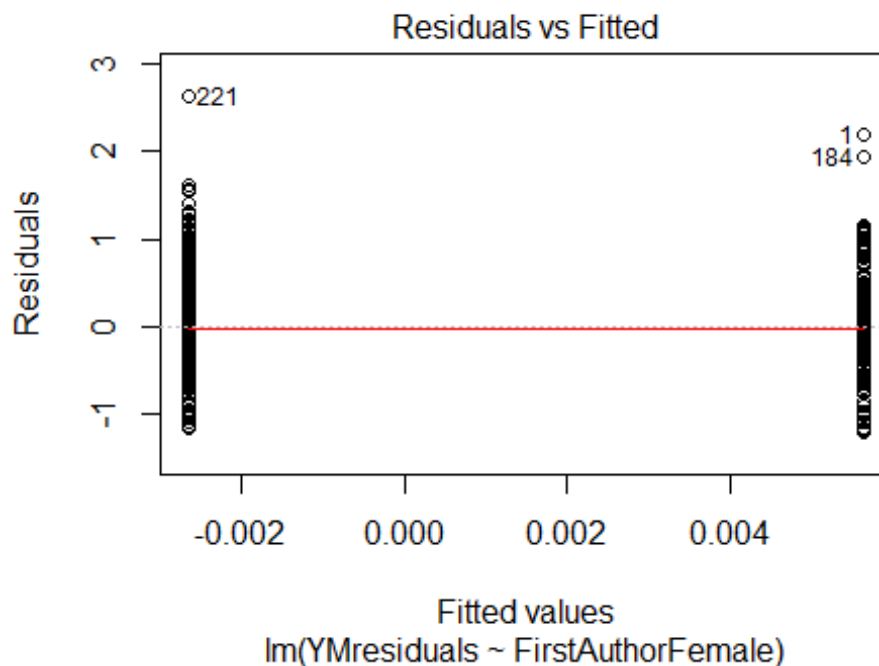
```
## 48 33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 47 48 26 55 39 32 40 39 26 26 28 38 40 41 53
## 2011 2012
## 46 31
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.45, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 2.37383915752614"
## [1] "Male first author team size 2018 geometric mean: 2.48441690882754"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.17501707105764"
## [1] "Male last author team size 2018 geometric mean: 3.00917403272406"

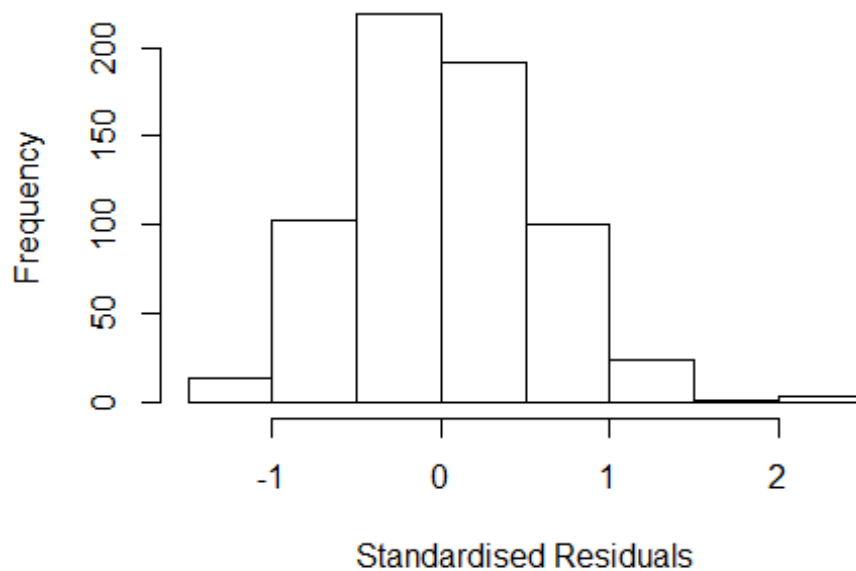
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 120, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.128	1	1.062
LastAuthorFemale	1.173	1	1.083
UniqueAuthors	1.692	4	1.068
Year	1.822	16	1.019

## Residuals from first and last author and team size



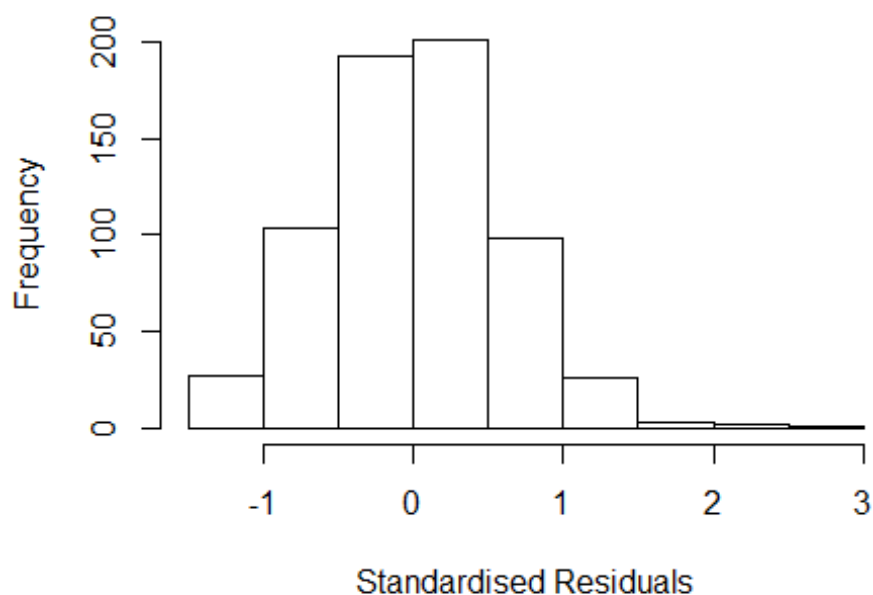
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2429 -0.3833 -0.0185 0.3904 2.4846
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6309 0.1121 5.63 2.7e-08 ***
## FirstAuthorFemale1 -0.0265 0.0493 -0.54 0.5910
## LastAuthorFemale1 0.0327 0.0499 0.66 0.5121
## UniqueAuthors2 0.2406 0.0565 4.26 2.4e-05 ***
## UniqueAuthors3 0.3073 0.0647 4.75 2.5e-06 ***
## UniqueAuthors4 0.4571 0.1093 4.18 3.3e-05 ***
## UniqueAuthors5 0.2991 0.1119 2.67 0.0077 **
## Year1997 -0.0231 0.1358 -0.17 0.8647
## Year1998 0.1453 0.1520 0.96 0.3393
## Year1999 -0.1800 0.1186 -1.52 0.1296
```

```

## Year2000          0.1912      0.1395      1.37      0.1708
## Year2001          0.0279      0.1645      0.17      0.8654
## Year2002          0.2199      0.1339      1.64      0.1011
## Year2003          0.2080      0.1313      1.58      0.1136
## Year2004          0.3534      0.1677      2.11      0.0355 *
## Year2005          0.2145      0.1535      1.40      0.1629
## Year2006          0.2285      0.1281      1.78      0.0749 .
## Year2007          0.2952      0.1264      2.34      0.0198 *
## Year2008          0.3873      0.1187      3.26      0.0012 **
## Year2009          0.2002      0.1490      1.34      0.1797
## Year2010          0.3312      0.1290      2.57      0.0104 *
## Year2011          0.2943      0.1297      2.27      0.0236 *
## Year2012          0.2590      0.1319      1.96      0.0501 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.568
## Multiple R-squared:  0.156, Adjusted R-squared:  0.127
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 612 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0167 0.8790 0.9520 0.9120 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.53e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.115 1 1.056
## LastAuthorFemale 1.125 1 1.060
## Year 1.113 16 1.003

```

## Residuals from first and last author

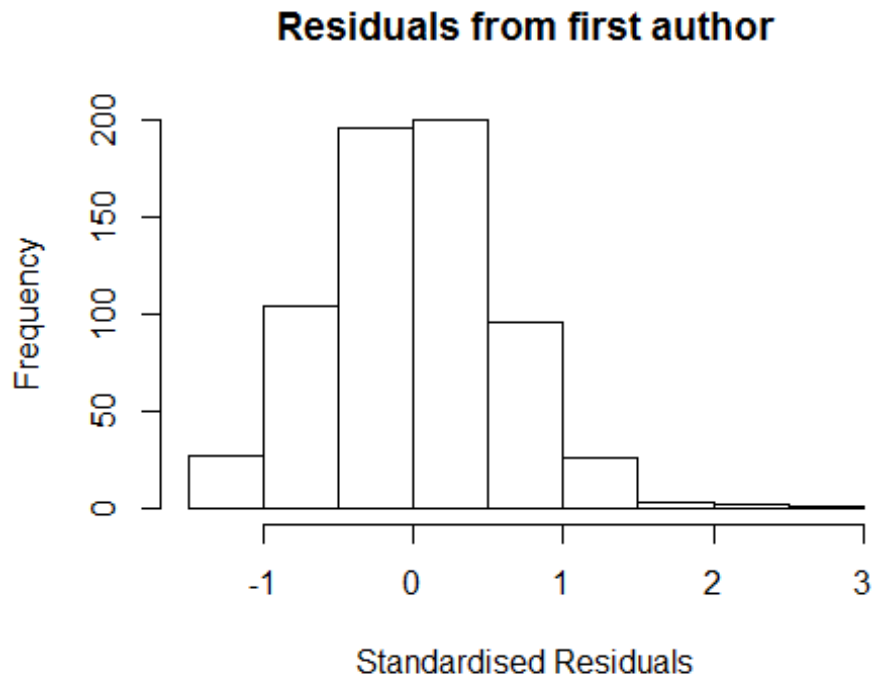


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 221 0033138567 3.249 1999      2906      3      2.691
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.18911 -0.38977  0.00107  0.40077  2.69107
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76439    0.12044   6.35 4.2e-10 ***
## FirstAuthorFemale1 -0.00235    0.05103  -0.05  0.96335
## LastAuthorFemale1  0.02404    0.05060   0.48  0.63493
## Year1997        -0.08147    0.14752  -0.55  0.58093
## Year1998         0.10297    0.16559   0.62  0.53430
## Year1999        -0.22815    0.12831  -1.78  0.07585 .
## Year2000         0.16475    0.15186   1.08  0.27839
## Year2001         0.01455    0.19633   0.07  0.94093
## Year2002         0.23847    0.14152   1.68  0.09248 .
## Year2003         0.22417    0.13388   1.67  0.09454 .
## Year2004         0.39479    0.16857   2.34  0.01949 *
## Year2005         0.25030    0.15773   1.59  0.11304
```

```

## Year2006          0.28682    0.13871    2.07  0.03906 *
## Year2007          0.36737    0.13259    2.77  0.00576 **
## Year2008          0.42472    0.12476    3.40  0.00071 ***
## Year2009          0.24455    0.15858    1.54  0.12354
## Year2010          0.37424    0.13495    2.77  0.00571 **
## Year2011          0.37845    0.13751    2.75  0.00609 **
## Year2012          0.32784    0.13957    2.35  0.01913 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.578
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0796
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 176 is an outlier with |weight| <= 0.00014 ( < 0.00015);
## 48 weights are ~= 1. The remaining 606 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0825 0.8690 0.9490 0.9100 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.53e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.057 1          1.028
## Year              1.057 16          1.002

```



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 221 0033138567 3.249 1999      2906      3      2.691
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.19981 -0.38981  0.00167  0.40370  2.69728
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.77577    0.11524   6.73 3.7e-11 ***
## FirstAuthorFemale1 0.00706    0.05042   0.14 0.88867
## Year1997      -0.08483    0.14675  -0.58 0.56342
## Year1998       0.10144    0.16536   0.61 0.53977
## Year1999      -0.23111    0.12784  -1.81 0.07111 .
## Year2000       0.16274    0.15145   1.07 0.28298
## Year2001       0.01259    0.19597   0.06 0.94881
## Year2002       0.23630    0.14153   1.67 0.09549 .
## Year2003       0.22064    0.13310   1.66 0.09787 .
## Year2004       0.39350    0.16802   2.34 0.01949 *
## Year2005       0.24862    0.15785   1.58 0.11575
## Year2006       0.28222    0.13835   2.04 0.04177 *
```



```

## Year2007          0.36330    0.13179    2.76  0.00601 **
## Year2008          0.42404    0.12475    3.40  0.00072 ***
## Year2009          0.24443    0.15855    1.54  0.12365
## Year2010          0.37408    0.13521    2.77  0.00583 **
## Year2011          0.37501    0.13715    2.73  0.00642 **
## Year2012          0.32831    0.13956    2.35  0.01895 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0808
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 176 is an outlier with |weight| <= 1.9e-05 ( < 0.00015);
## 53 weights are ~= 1. The remaining 601 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0854 0.8710 0.9480 0.9090 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.53e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.07 1          1.035
## Year          1.07 16          1.002
##
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 221 0033138567 3.249 1999          2906          3          2.691
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min          1Q      Median          3Q          Max

```

```

## -1.18807 -0.38862 0.00127 0.40112 2.69067
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7631    0.1171     6.52 1.4e-10 ***
## LastAuthorFemale1 0.0232    0.0498     0.47 0.6421
## Year1997       -0.0813    0.1474    -0.55 0.5812
## Year1998        0.1035    0.1654     0.63 0.5316
## Year1999       -0.2279    0.1284    -1.77 0.0765 .
## Year2000        0.1651    0.1519     1.09 0.2778
## Year2001        0.0145    0.1961     0.07 0.9409
## Year2002        0.2389    0.1416     1.69 0.0921 .
## Year2003        0.2244    0.1340     1.68 0.0944 .
## Year2004        0.3950    0.1687     2.34 0.0196 *
## Year2005        0.2504    0.1578     1.59 0.1131
## Year2006        0.2870    0.1388     2.07 0.0390 *
## Year2007        0.3676    0.1327     2.77 0.0058 **
## Year2008        0.4250    0.1248     3.40 0.0007 ***
## Year2009        0.2451    0.1589     1.54 0.1236
## Year2010        0.3744    0.1350     2.77 0.0057 **
## Year2011        0.3787    0.1375     2.75 0.0061 **
## Year2012        0.3281    0.1396     2.35 0.0191 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.577
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0813
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 176 is an outlier with |weight| <= 6.7e-05 ( < 0.00015);
## 47 weights are ~= 1. The remaining 607 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0804 0.8680 0.9490 0.9100 0.9850 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.53e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
## trace.lev    mts    compute.rd
##           0         1000         0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)

```

```

## [1] "Sample size for the above analysis: 655"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2920"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011
##    3    8    1    7    7    4    1    1    2    3    2    3   11    6    1
## 2012
##    1
##
## 1996 1997 1998 1999 2000 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011
##    3    4    1    5    2    3    1    1    2    3    2    3    9    5    1
## 2012
##    1
##
## 1996 1997 1998 1999 2000 2001 2002 2004 2005 2006 2007 2008 2009 2010 2011
##    2    3    1    4    2    3    1    1    2    3    2    2    9    5    1
## 2012
##    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## [1] "Regression 3: First author gender, Year as factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

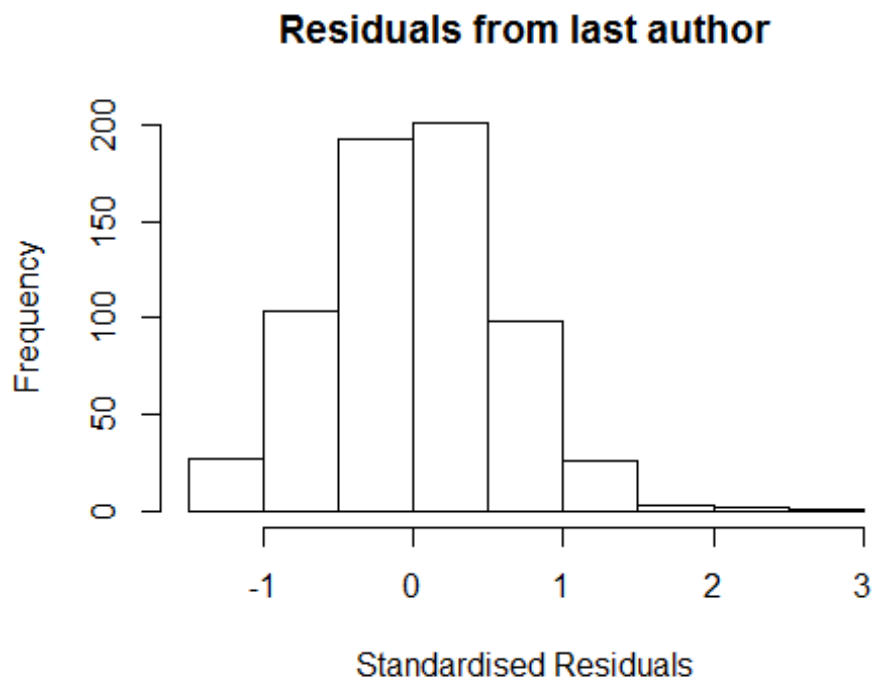
```

```
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## [1] "Regression 4: Last author gender, Year as factors"

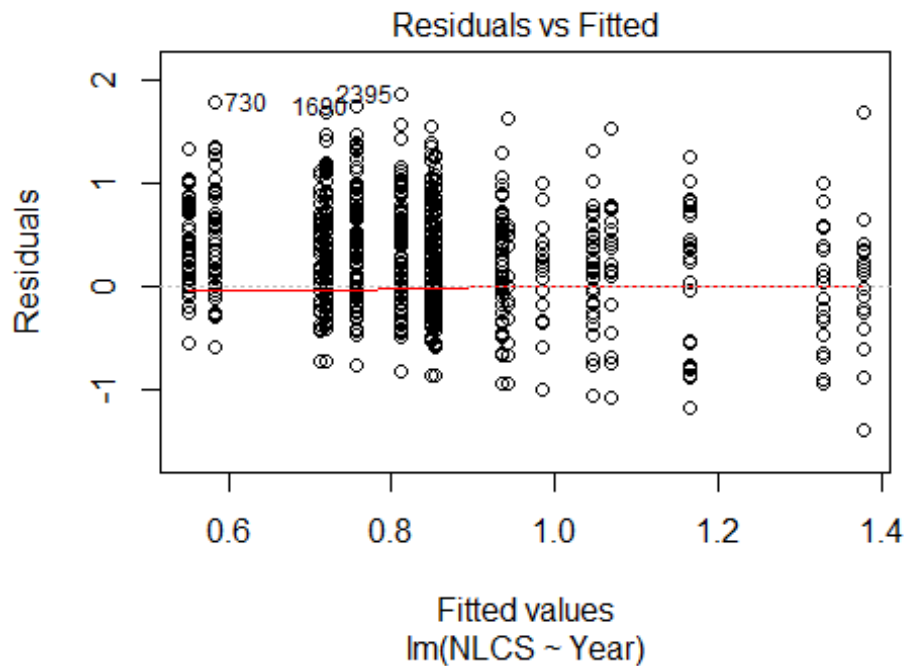
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations
```

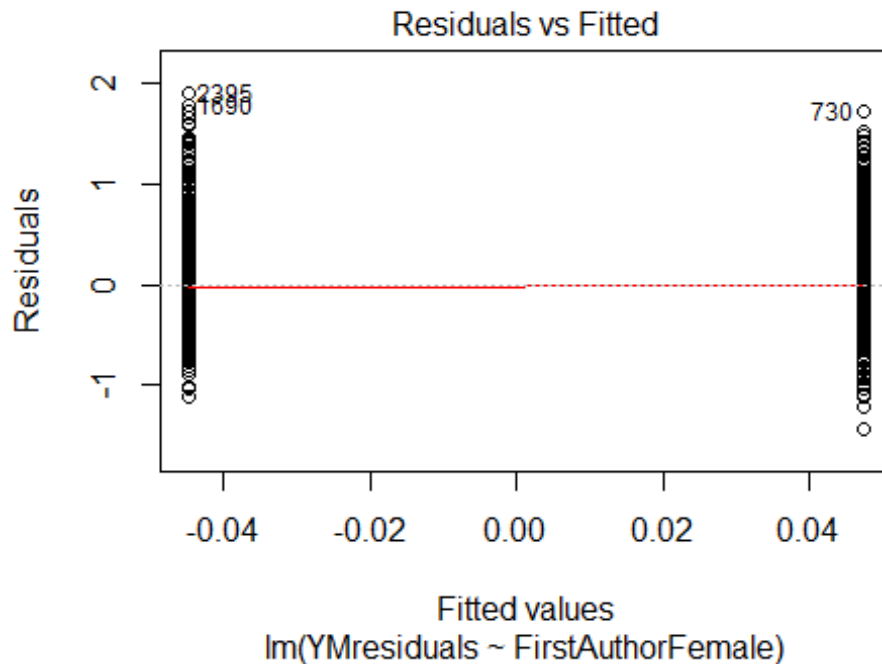


```
## [1] "Sample size for the above analysis: 42"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2921"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   51   70   76   72   92   99  103   86  163  195  122  126  153  199  207
## 2011 2012
##  215  215
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   21   23   21   28   26   52   41  125  142   76   78  103  142  142
```

```
## 2011 2012
## 156 158
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 22 21 21 21 27 25 45 35 117 131 68 72 96 137 132
## 2011 2012
## 140 149
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 31, df = 16, p-value = 0.01
```

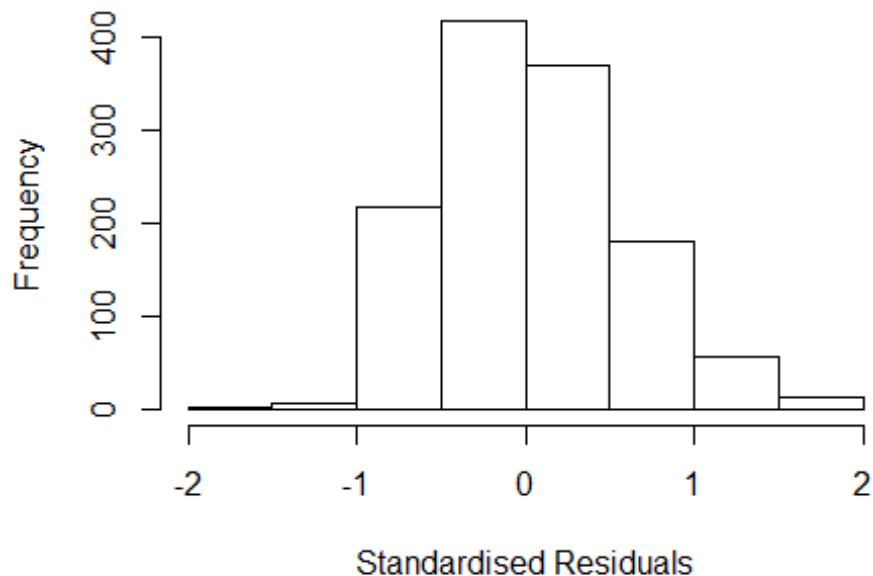


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.8, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 2.55585763938264"
## [1] "Male first author team size 2018 geometric mean: 2.76679385330417"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2400, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.34160006302499"
## [1] "Male last author team size 2018 geometric mean: 2.94959086766963"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.514 1      1.230
## LastAuthorFemale  1.463 1      1.210
## UniqueAuthors    1.574 4      1.058
## Year              1.604 16     1.015
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5262 -0.4136 -0.0223 0.3807 1.8965
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0394 0.1375 7.56 7.9e-14 ***
## FirstAuthorFemale1 0.0142 0.0390 0.36 0.71625
## LastAuthorFemale1 0.0726 0.0387 1.88 0.06093 .
## UniqueAuthors2 0.3264 0.0425 7.67 3.4e-14 ***
## UniqueAuthors3 0.4000 0.0504 7.93 4.9e-15 ***
## UniqueAuthors4 0.3744 0.0756 4.95 8.3e-07 ***
## UniqueAuthors5 0.5060 0.0753 6.72 2.7e-11 ***
## Year1997 0.0725 0.1882 0.39 0.70021
## Year1998 -0.2820 0.1768 -1.59 0.11108
## Year1999 -0.3749 0.1702 -2.20 0.02776 *
```

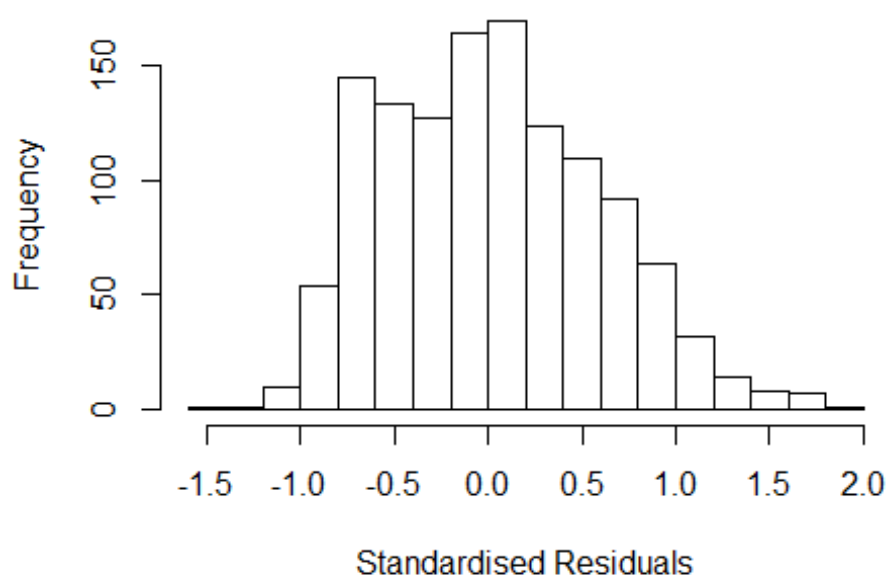
```

## Year2000          -0.1788      0.1828   -0.98   0.32818
## Year2001          -0.1819      0.2095   -0.87   0.38562
## Year2002          -0.4508      0.1660   -2.72   0.00670 **
## Year2003          -0.3368      0.1675   -2.01   0.04453 *
## Year2004          -0.6930      0.1420   -4.88   1.2e-06 ***
## Year2005          -0.6992      0.1406   -4.97   7.5e-07 ***
## Year2006          -0.5941      0.1461   -4.07   5.1e-05 ***
## Year2007          -0.3967      0.1499   -2.65   0.00825 **
## Year2008          -0.4537      0.1476   -3.07   0.00216 **
## Year2009          -0.6028      0.1473   -4.09   4.5e-05 ***
## Year2010          -0.5214      0.1475   -3.53   0.00042 ***
## Year2011          -0.6611      0.1441   -4.59   4.9e-06 ***
## Year2012          -0.5764      0.1451   -3.97   7.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.56
## Multiple R-squared:  0.194, Adjusted R-squared:  0.18
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 104 weights are ~= 1. The remaining 1155 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.228  0.878  0.947   0.910   0.983   0.999
## Algorithmic parameters:
##           tuning.chi              bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           7.94e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.387 1           1.178
## LastAuthorFemale  1.357 1           1.165
## Year              1.141 16           1.004

```



## Residuals from first and last author



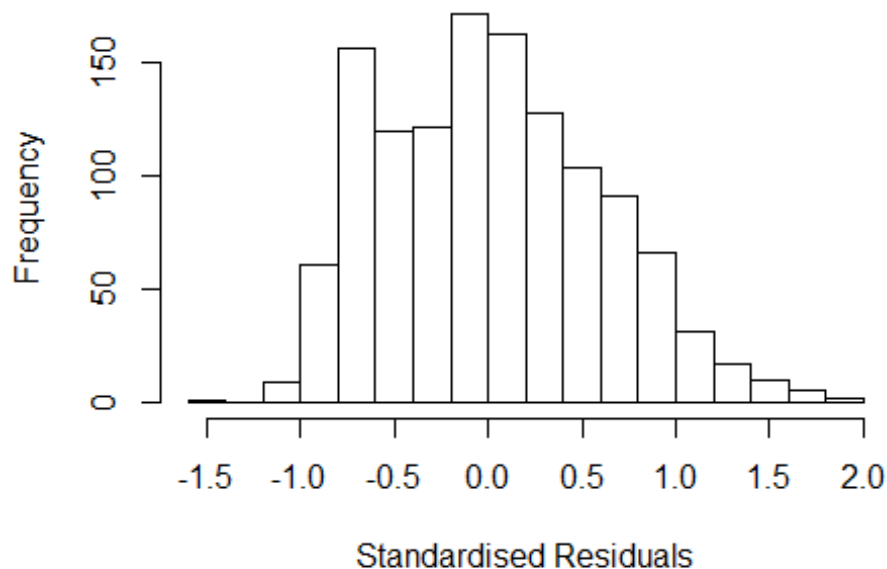
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4269 -0.4391 -0.0131 0.4201 1.8630
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2770 0.1219 10.47 < 2e-16 ***
## FirstAuthorFemale1 0.0556 0.0398 1.40 0.16220
## LastAuthorFemale1 0.0943 0.0397 2.37 0.01774 *
## Year1997 -0.0236 0.1739 -0.14 0.89216
## Year1998 -0.3757 0.1700 -2.21 0.02728 *
## Year1999 -0.4252 0.1616 -2.63 0.00861 **
## Year2000 -0.2517 0.1821 -1.38 0.16713
## Year2001 -0.2073 0.2215 -0.94 0.34943
## Year2002 -0.5050 0.1559 -3.24 0.00123 **
## Year2003 -0.3529 0.1559 -2.26 0.02381 *
## Year2004 -0.8211 0.1277 -6.43 1.8e-10 ***
## Year2005 -0.8592 0.1263 -6.80 1.6e-11 ***
```

```

## Year2006          -0.6860      0.1339   -5.12  3.5e-07 ***
## Year2007          -0.4176      0.1394   -3.00  0.00279 **
## Year2008          -0.4998      0.1347   -3.71  0.00021 ***
## Year2009          -0.6514      0.1360   -4.79  1.9e-06 ***
## Year2010          -0.5396      0.1352   -3.99  7.0e-05 ***
## Year2011          -0.6866      0.1313   -5.23  2.0e-07 ***
## Year2012          -0.5703      0.1322   -4.31  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.111, Adjusted R-squared:  0.0978
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1153 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.313  0.867  0.946  0.913  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.102 1          1.050
## Year              1.102 16          1.003

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.41177 -0.43942 -0.00877 0.42270 1.93147
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3108 0.1158 11.32 < 2e-16 ***
## FirstAuthorFemale1 0.1010 0.0355 2.85 0.00450 **
## Year1997 -0.0370 0.1686 -0.22 0.82640
## Year1998 -0.3779 0.1677 -2.25 0.02440 *
## Year1999 -0.4466 0.1585 -2.82 0.00491 **
## Year2000 -0.2640 0.1804 -1.46 0.14351
## Year2001 -0.2261 0.2182 -1.04 0.30028
## Year2002 -0.5146 0.1527 -3.37 0.00078 ***
## Year2003 -0.3584 0.1522 -2.36 0.01867 *
## Year2004 -0.8400 0.1226 -6.85 1.1e-11 ***
## Year2005 -0.8713 0.1214 -7.17 1.2e-12 ***
## Year2006 -0.6999 0.1296 -5.40 8.1e-08 ***
```

```

## Year2007          -0.4294      0.1349   -3.18  0.00150 **
## Year2008          -0.5167      0.1302   -3.97  7.7e-05 ***
## Year2009          -0.6583      0.1319   -4.99  6.8e-07 ***
## Year2010          -0.5525      0.1307   -4.23  2.5e-05 ***
## Year2011          -0.6979      0.1266   -5.51  4.3e-08 ***
## Year2012          -0.5782      0.1279   -4.52  6.8e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.599
## Multiple R-squared:  0.107, Adjusted R-squared:  0.0945
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 120 weights are ~= 1. The remaining 1139 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.277  0.865  0.943  0.912  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      7.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.077 1      1.038
## Year      1.077 16      1.002

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.4112 -0.4291 -0.0241  0.4174  1.8191

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2905     0.1229   10.50 < 2e-16 ***
## LastAuthorFemale1 0.1206     0.0353    3.41 0.00066 ***
## Year1997         -0.0188     0.1754   -0.11 0.91467
## Year1998         -0.3828     0.1706   -2.24 0.02499 *
## Year1999         -0.4219     0.1618   -2.61 0.00921 **
## Year2000         -0.2585     0.1818   -1.42 0.15537
## Year2001         -0.1950     0.2231   -0.87 0.38215
## Year2002         -0.5041     0.1567   -3.22 0.00133 **
## Year2003         -0.3548     0.1568   -2.26 0.02378 *
## Year2004         -0.8258     0.1285   -6.43 1.9e-10 ***
## Year2005         -0.8614     0.1272   -6.77 1.9e-11 ***
## Year2006         -0.6905     0.1348   -5.12 3.5e-07 ***
## Year2007         -0.4141     0.1405   -2.95 0.00327 **
## Year2008         -0.4969     0.1356   -3.67 0.00026 ***
## Year2009         -0.6530     0.1367   -4.78 2.0e-06 ***
## Year2010         -0.5351     0.1360   -3.93 8.8e-05 ***
## Year2011         -0.6828     0.1321   -5.17 2.8e-07 ***
## Year2012         -0.5663     0.1329   -4.26 2.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
## Multiple R-squared:  0.109, Adjusted R-squared:  0.0971
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 107 weights are ~= 1. The remaining 1152 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.339 0.871 0.946 0.913 0.984 0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      7.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1259"
## [1] ""

```

```

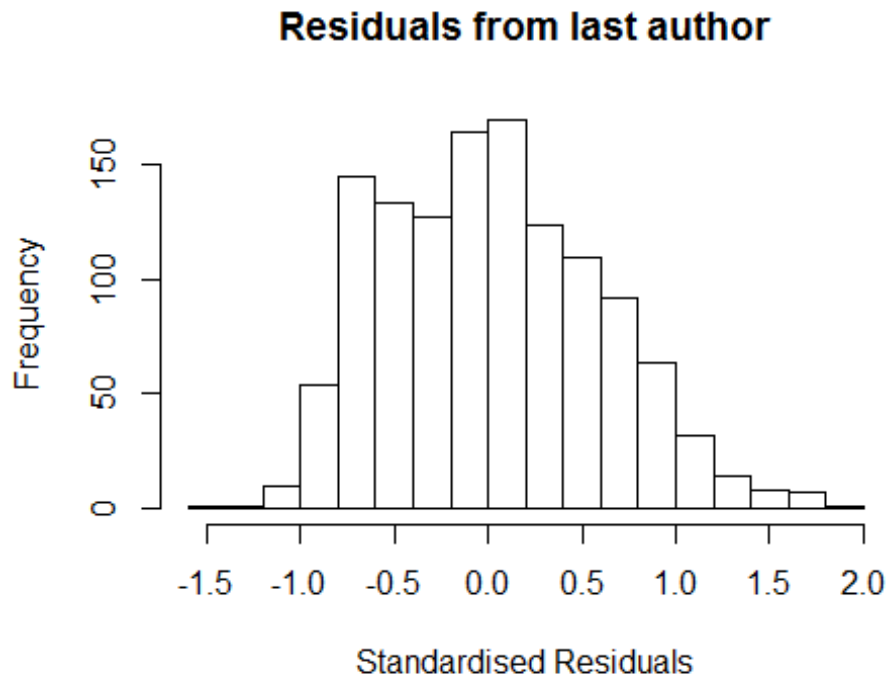
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2922"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 36 38 35 3 4 7 16 11 29 38 44 27 49 48
## 2011 2012
## 29 23
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 30 34 33 33 1 3 6 16 8 27 32 42 25 44 43
## 2011 2012
## 27 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 29 31 33 30 1 3 5 16 8 27 31 40 23 44 42
## 2011 2012
## 27 20
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.38799247449692"
## [1] "Male first author team size 2018 geometric mean: 1.57605860014924"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 66, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.29723000039817"
## [1] "Male last author team size 2018 geometric mean: 1.58583317513724"

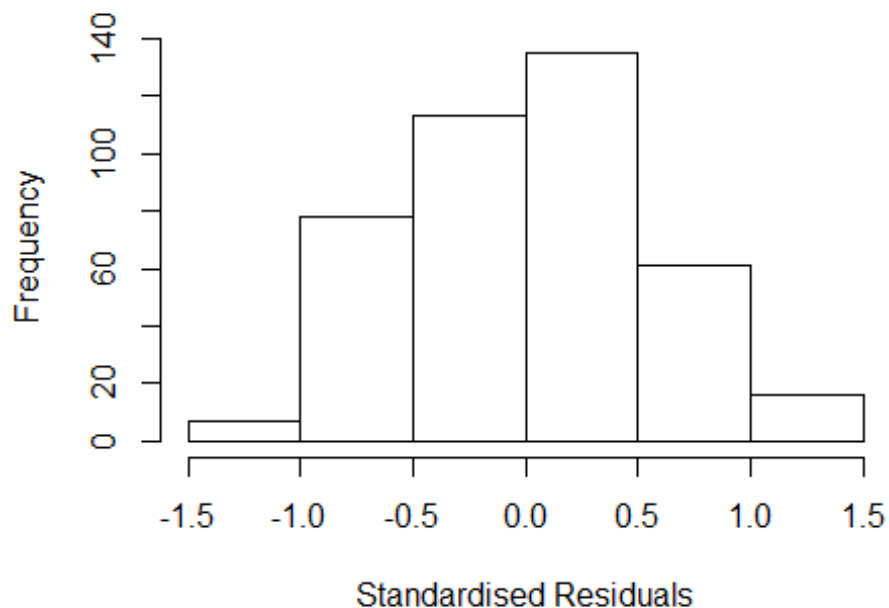
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 62, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.604 1          1.614
## LastAuthorFemale  2.593 1          1.610
## UniqueAuthors    4.641 4          1.212
## Year              8.927 16          1.071
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4122 -0.4263 0.0234 0.4199 1.4538
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96236 0.09979 9.64 < 2e-16 ***
## FirstAuthorFemale1 -0.06919 0.07960 -0.87 0.38527
## LastAuthorFemale1 0.09535 0.07710 1.24 0.21696
## UniqueAuthors2 0.28119 0.08553 3.29 0.00110 **
## UniqueAuthors3 0.24162 0.09539 2.53 0.01170 *
## UniqueAuthors4 0.36752 0.10588 3.47 0.00058 ***
## UniqueAuthors5 0.53381 0.10472 5.10 5.4e-07 ***
## Year1997 -0.18890 0.14566 -1.30 0.19545
## Year1998 -0.14671 0.15643 -0.94 0.34891
## Year1999 -0.05843 0.13303 -0.44 0.66074
```

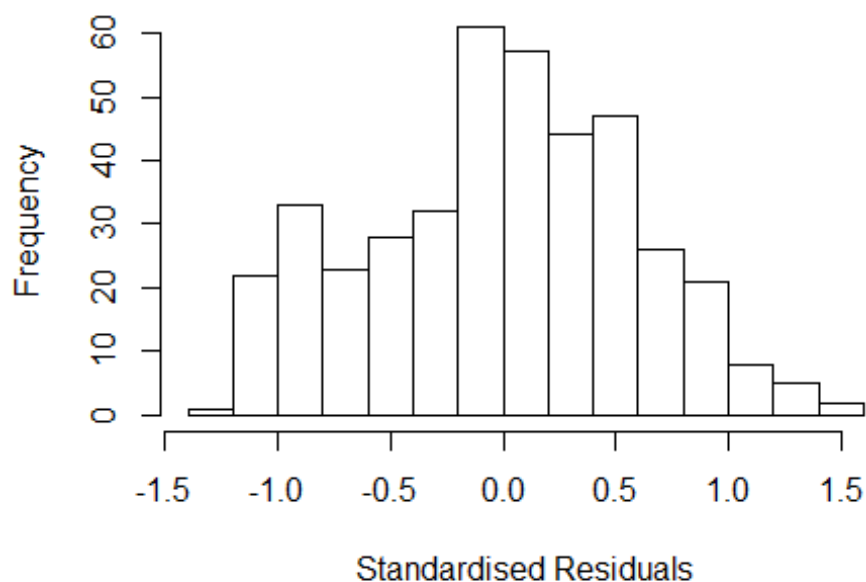


```

## Year2000      -0.98852    0.09572   -10.33   < 2e-16 ***
## Year2001      -0.37321    0.20145    -1.85   0.06470 .
## Year2002      -0.35619    0.19842    -1.80   0.07341 .
## Year2003       0.02446    0.23793     0.10   0.91816
## Year2004      -0.36506    0.15576    -2.34   0.01959 *
## Year2005      -0.18718    0.15834    -1.18   0.23789
## Year2006       0.03525    0.14719     0.24   0.81085
## Year2007      -0.15178    0.13614    -1.11   0.26559
## Year2008      -0.00281    0.19290     -0.01   0.98837
## Year2009      -0.31320    0.12789    -2.45   0.01477 *
## Year2010      -0.19198    0.12248    -1.57   0.11783
## Year2011       0.16861    0.12289     1.37   0.17085
## Year2012      -0.30007    0.17023    -1.76   0.07873 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.604
## Multiple R-squared:  0.119, Adjusted R-squared:  0.0689
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 376 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.542  0.859  0.952  0.916  0.983  0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           2.44e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
## trace.lev    mts    compute.rd
##           0         1000         0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.820 1         1.679
## LastAuthorFemale 2.746 1         1.657
## Year              2.545 16         1.030

```

## Residuals from first and last author



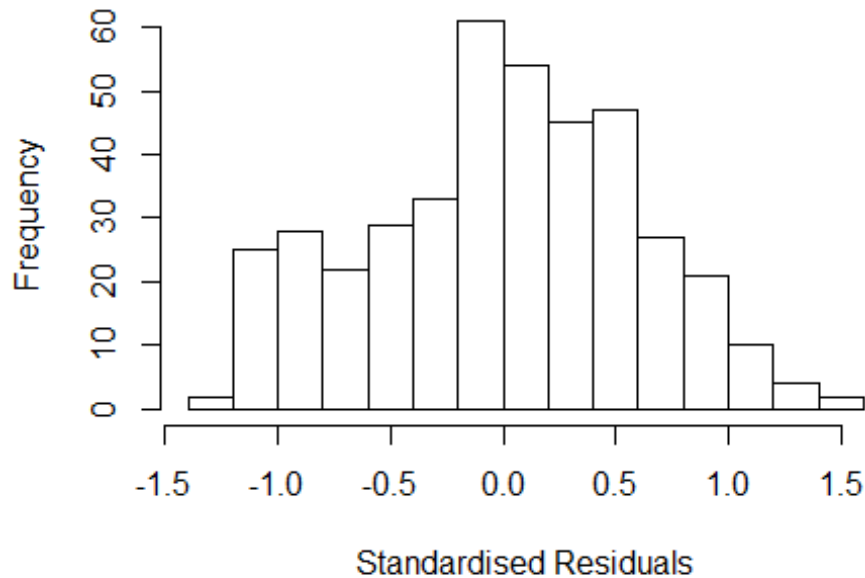
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2647 -0.4277 0.0338 0.4204 1.4931
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00959 0.10208 9.89 <2e-16 ***
## FirstAuthorFemale1 -0.03197 0.08681 -0.37 0.713
## LastAuthorFemale1 0.10675 0.08211 1.30 0.194
## Year1997 -0.05486 0.14434 -0.38 0.704
## Year1998 -0.07496 0.16634 -0.45 0.653
## Year1999 0.02325 0.13112 0.18 0.859
## Year2000 -1.08437 0.09574 -11.33 <2e-16 ***
## Year2001 -0.38858 0.14325 -2.71 0.007 **
## Year2002 -0.26191 0.20713 -1.26 0.207
## Year2003 -0.01900 0.23166 -0.08 0.935
## Year2004 -0.31006 0.19527 -1.59 0.113
## Year2005 -0.14261 0.16616 -0.86 0.391
```

```

## Year2006          0.05840      0.14172      0.41      0.680
## Year2007          -0.00215      0.14006     -0.02      0.988
## Year2008           0.01214      0.20782      0.06      0.953
## Year2009          -0.26566      0.13410     -1.98      0.048 *
## Year2010          -0.16969      0.13292     -1.28      0.203
## Year2011           0.18029      0.12436      1.45      0.148
## Year2012          -0.17715      0.16667     -1.06      0.288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0523, Adjusted R-squared:  0.00866
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 373 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.536  0.858  0.948  0.913  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.806 1      1.344
## Year              1.806 16      1.019

```

## Residuals from first author



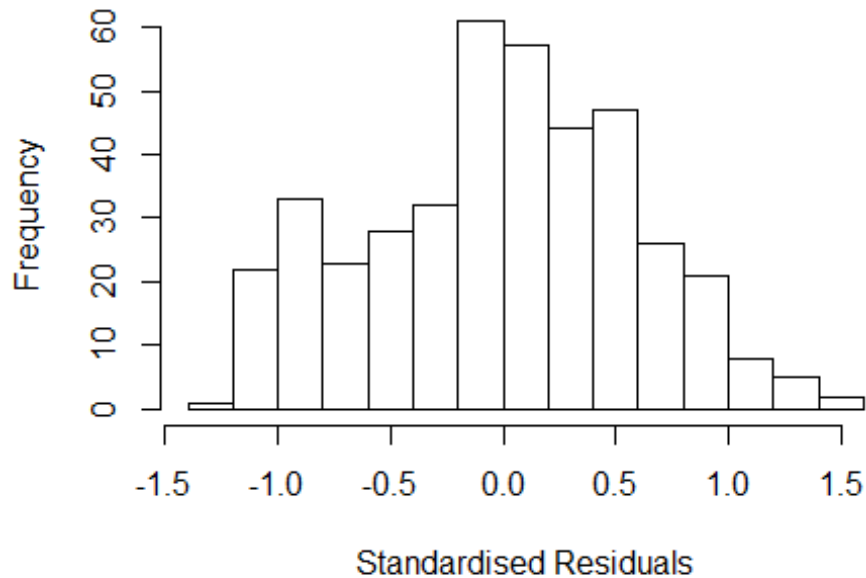
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2417 -0.4138 0.0149 0.4381 1.4659
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03075 0.09967 10.34 <2e-16 ***
## FirstAuthorFemale1 0.04001 0.07062 0.57 0.571
## Year1997 -0.05774 0.14350 -0.40 0.688
## Year1998 -0.06988 0.16534 -0.42 0.673
## Year1999 0.02906 0.13117 0.22 0.825
## Year2000 -1.07076 0.09424 -11.36 <2e-16 ***
## Year2001 -0.40955 0.16594 -2.47 0.014 *
## Year2002 -0.29693 0.20170 -1.47 0.142
## Year2003 -0.02193 0.23252 -0.09 0.925
## Year2004 -0.32757 0.19047 -1.72 0.086 .
## Year2005 -0.14416 0.16646 -0.87 0.387
## Year2006 0.03992 0.14083 0.28 0.777
```

```

## Year2007          0.00173    0.13966    0.01    0.990
## Year2008          0.01979    0.20766    0.10    0.924
## Year2009         -0.26598    0.13241   -2.01    0.045 *
## Year2010         -0.16367    0.13262   -1.23    0.218
## Year2011          0.17093    0.12263    1.39    0.164
## Year2012         -0.17257    0.16793   -1.03    0.305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.628
## Multiple R-squared:  0.0478, Adjusted R-squared:  0.00654
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 43 weights are ~= 1. The remaining 367 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.565  0.862  0.947   0.914  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.809 1          1.345
## Year            1.809 16          1.019

```

## Residuals from last author



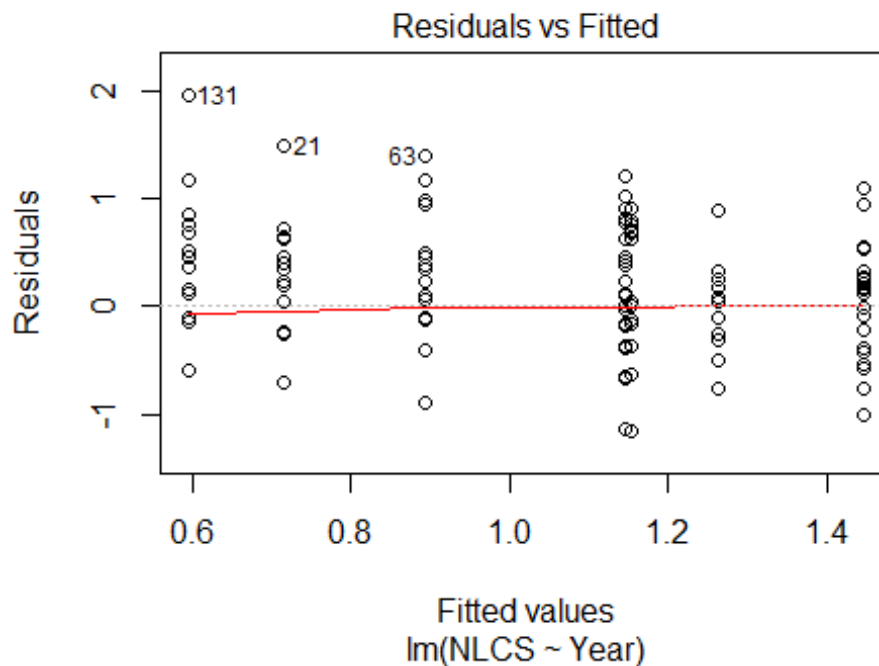
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2667 -0.4326 0.0359 0.4229 1.4999
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00271 0.09945 10.08 <2e-16 ***
## LastAuthorFemale1 0.08560 0.06666 1.28 0.1999
## Year1997 -0.05517 0.14384 -0.38 0.7015
## Year1998 -0.07307 0.16472 -0.44 0.6576
## Year1999 0.02214 0.13079 0.17 0.8657
## Year2000 -1.08831 0.09460 -11.50 <2e-16 ***
## Year2001 -0.39942 0.14481 -2.76 0.0061 **
## Year2002 -0.27589 0.20115 -1.37 0.1710
## Year2003 -0.01969 0.22966 -0.09 0.9317
## Year2004 -0.31022 0.19282 -1.61 0.1085
## Year2005 -0.14231 0.16581 -0.86 0.3912
## Year2006 0.05192 0.14023 0.37 0.7114
```

```

## Year2007          -0.00247      0.13973    -0.02    0.9859
## Year2008           0.01066      0.20687     0.05    0.9589
## Year2009          -0.26474      0.13382    -1.98    0.0486 *
## Year2010          -0.16962      0.13296    -1.28    0.2028
## Year2011           0.17844      0.12410     1.44    0.1513
## Year2012          -0.17355      0.16651    -1.04    0.2979
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.623
## Multiple R-squared:  0.0515, Adjusted R-squared:  0.0104
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 373 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.541  0.860  0.950  0.915  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.44e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 410"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 2923"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 2006 2007 2008 2009 2010 2011 2012
##   29   25   33   37   32   20   38
##
## 2006 2007 2008 2009 2010 2011 2012
##   29   19   32   33   23   13   36
##
## 2006 2007 2008 2009 2010 2011 2012
##   27   18   28   32   22   13   31

```

```
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 4.6, df = 6, p-value = 0.6
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.11, df = 1, p-value = 0.7

## [1] "Female first author team size 2018 geometric mean: 2.70519930996487"
## [1] "Male first author team size 2018 geometric mean: 2.28223948067992"

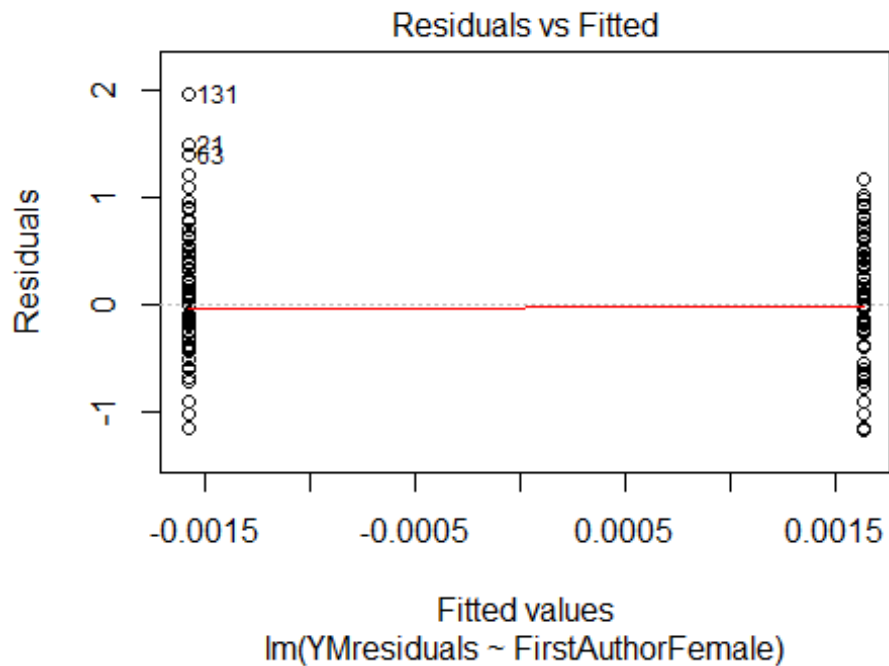
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 85, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
```



```
## [1] "Female last author team size 2018 geometric mean: 2.69856569534713"
## [1] "Male last author team size 2018 geometric mean: 2.34245432167617"

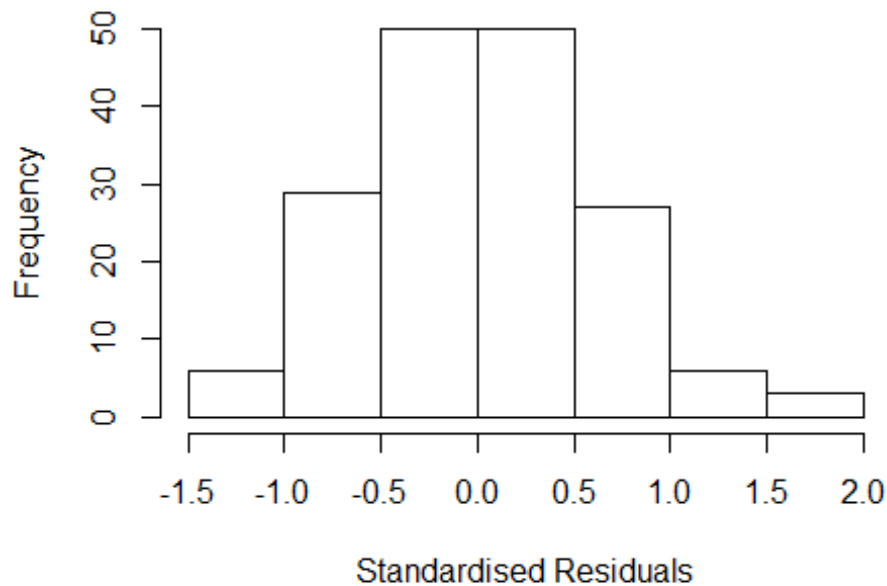
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 80, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.375	1	1.172
LastAuthorFemale	1.453	1	1.205
UniqueAuthors	2.474	4	1.120
Year	2.881	6	1.092

## Residuals from first and last author and team size



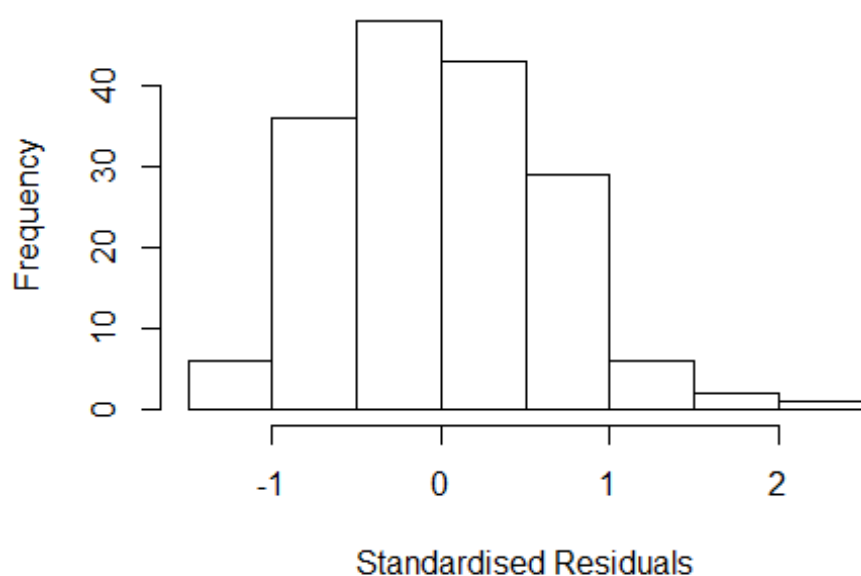
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.22741 -0.40600 0.00651 0.40943 1.86148
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.65718 0.13836 4.75 4.5e-06 ***
## FirstAuthorFemale1 -0.02829 0.10479 -0.27 0.7876
## LastAuthorFemale1 -0.03776 0.10265 -0.37 0.7135
## UniqueAuthors2 0.25476 0.13613 1.87 0.0631 .
## UniqueAuthors3 0.15018 0.11927 1.26 0.2098
## UniqueAuthors4 0.00474 0.16769 0.03 0.9775
## UniqueAuthors5 0.29176 0.23326 1.25 0.2129
## Year2007 0.57023 0.21052 2.71 0.0075 **
## Year2008 0.13240 0.16868 0.78 0.4337
## Year2009 -0.18513 0.14865 -1.25 0.2148
```

```

## Year2010          0.79697    0.16434    4.85  2.9e-06 ***
## Year2011          0.51405    0.14860    3.46  0.0007 ***
## Year2012          0.26432    0.18111    1.46  0.1464
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.597
## Multiple R-squared:  0.245, Adjusted R-squared:  0.187
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 148 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.311  0.871  0.944  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.253 1 1.119
## LastAuthorFemale 1.388 1 1.178
## Year 1.217 6 1.017

```

## Residuals from first and last author



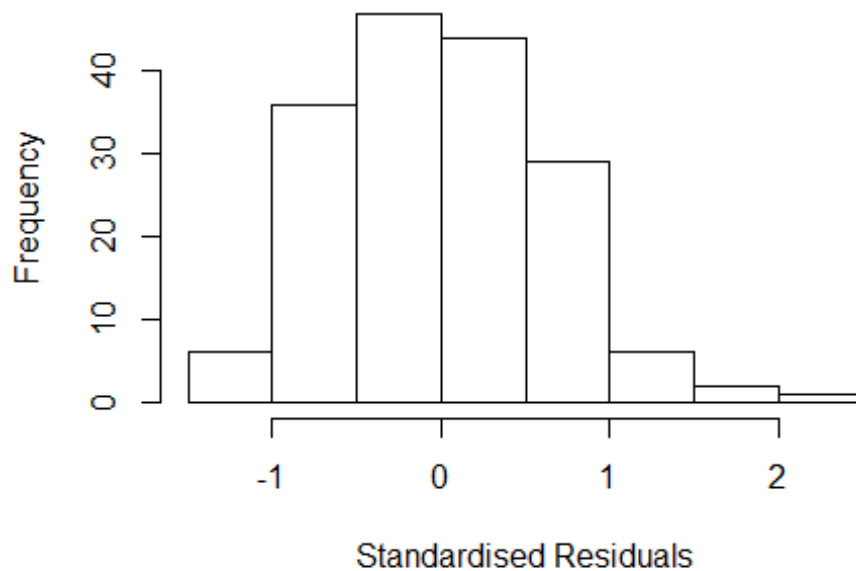
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2502 -0.4780 -0.0206 0.4405 2.0410
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7323 0.1305 5.61 8.5e-08 ***
## FirstAuthorFemale1 -0.0460 0.1046 -0.44 0.66076
## LastAuthorFemale1 -0.0173 0.1047 -0.17 0.86873
## Year2007 0.5179 0.2185 2.37 0.01893 *
## Year2008 0.1312 0.1768 0.74 0.45902
## Year2009 -0.1673 0.1548 -1.08 0.28124
## Year2010 0.7859 0.1666 4.72 5.1e-06 ***
## Year2011 0.5654 0.1579 3.58 0.00045 ***
## Year2012 0.3866 0.1588 2.43 0.01603 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## Robust residual standard error: 0.606
## Multiple R-squared: 0.221, Adjusted R-squared: 0.182
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 158 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.234 0.879 0.939 0.911 0.989 0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.071 1      1.035
## Year              1.071 6      1.006

```

## Residuals from first author



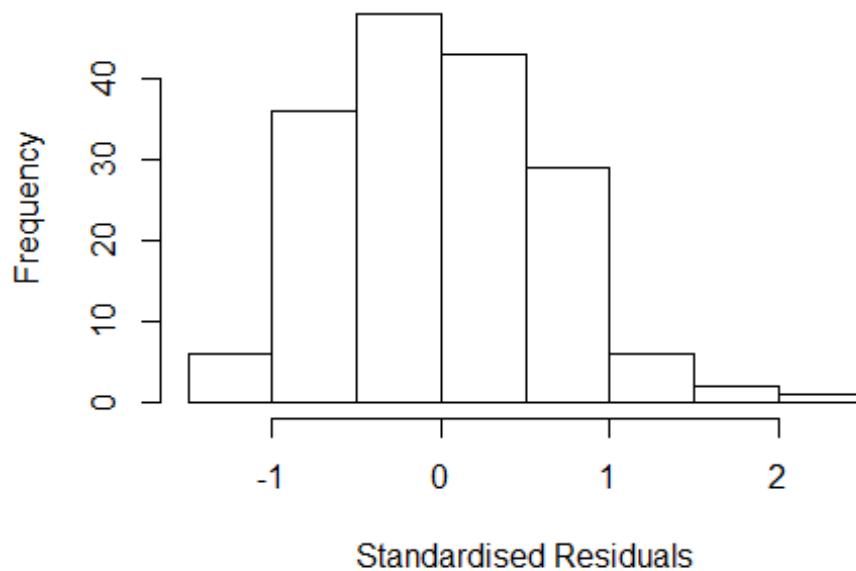
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2472 -0.4713 -0.0224 0.4360 2.0566
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7281 0.1257 5.79 3.5e-08 ***
## FirstAuthorFemale1 -0.0561 0.0964 -0.58 0.56162
## Year2007 0.5190 0.2228 2.33 0.02104 *
## Year2008 0.1305 0.1773 0.74 0.46262
## Year2009 -0.1687 0.1540 -1.10 0.27477
## Year2010 0.7888 0.1675 4.71 5.3e-06 ***
## Year2011 0.5693 0.1593 3.57 0.00047 ***
## Year2012 0.3872 0.1577 2.46 0.01512 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
```

```

## Multiple R-squared:  0.222, Adjusted R-squared:  0.188
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 13 weights are ~= 1. The remaining 158 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.218  0.875  0.937   0.910   0.988   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           5.85e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
##   nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##           500          50         2         1         1000         200
##   trace.lev      mts    compute.rd
##           0          1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.157 1         1.075
## Year             1.157 6         1.012

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2394 -0.4809 -0.0313 0.4621 2.0060
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7217 0.1281 5.64 7.5e-08 ***
## LastAuthorFemale1 -0.0417 0.0972 -0.43 0.66845
## Year2007 0.5177 0.2218 2.33 0.02079 *
## Year2008 0.1293 0.1778 0.73 0.46810
## Year2009 -0.1677 0.1547 -1.08 0.28016
## Year2010 0.7770 0.1650 4.71 5.3e-06 ***
## Year2011 0.5541 0.1581 3.51 0.00059 ***
## Year2012 0.3908 0.1592 2.46 0.01513 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.601
```

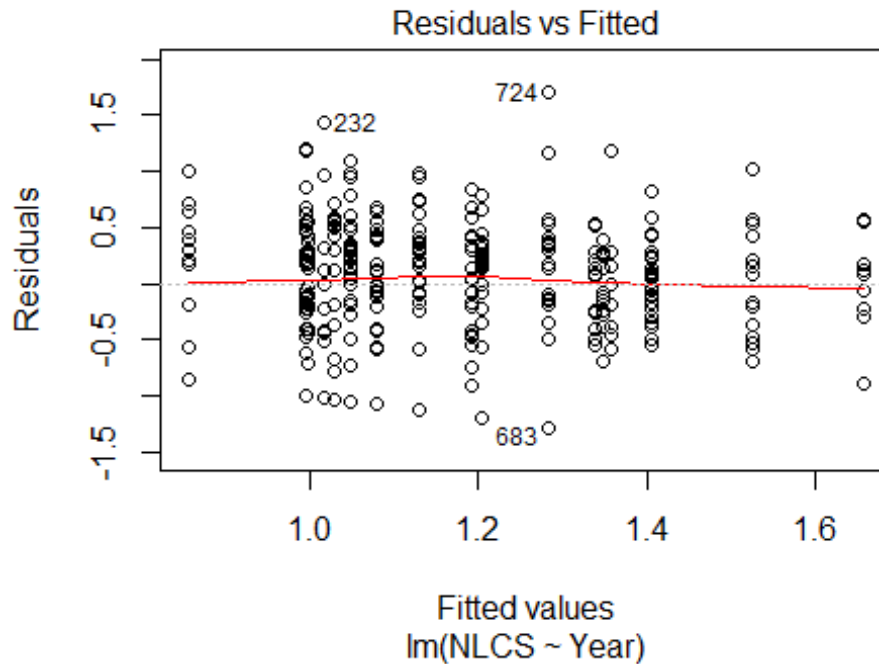


```

## Multiple R-squared:  0.22,   Adjusted R-squared:  0.187
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 159 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.243  0.873  0.938   0.911  0.988   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           5.85e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it    best.r.s    k.fast.s      k.max maxit.scale
##           500          50         2         1         1000         200
## trace.lev      mts    compute.rd
##           0          1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 171"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3000"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   44  27  49  46  51  47  48  35  44  22  26  40  45  19  29
## 2011 2012
##   42  38
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   9  14  24  11  15  27  13  21  15  19  26  31  18  25
## 2011 2012
##   35  33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    7   7  14  20  10  15  26   9  20  13  17  23  30  18  21
## 2011 2012
##   33  30
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances

```

```
##
## data: NLCS by Year
## Bartlett's K-squared = 41, df = 16, p-value = 6e-04
```



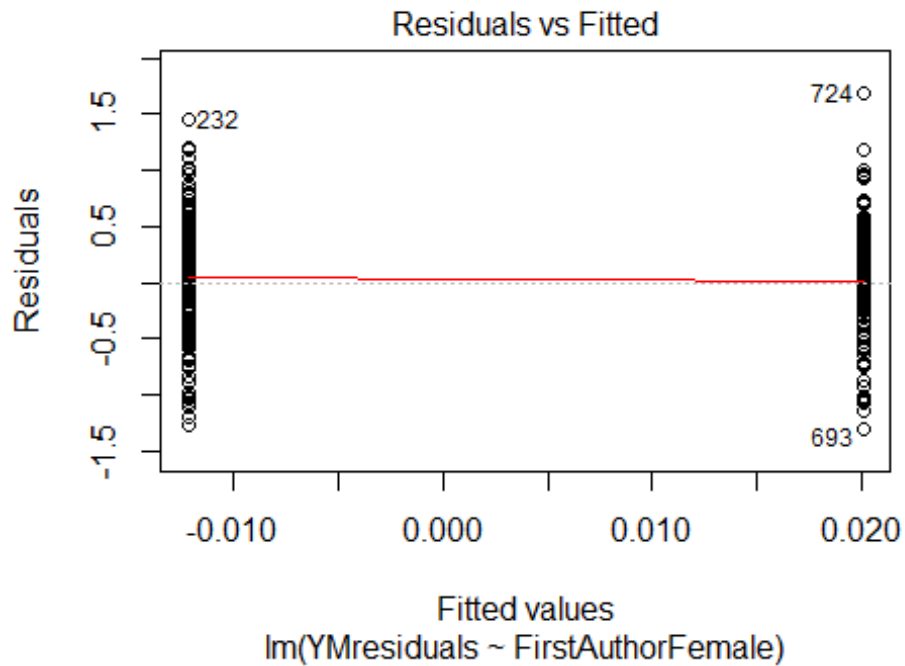
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.8, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 4.57590035167661"
## [1] "Male first author team size 2018 geometric mean: 2.94310442781958"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

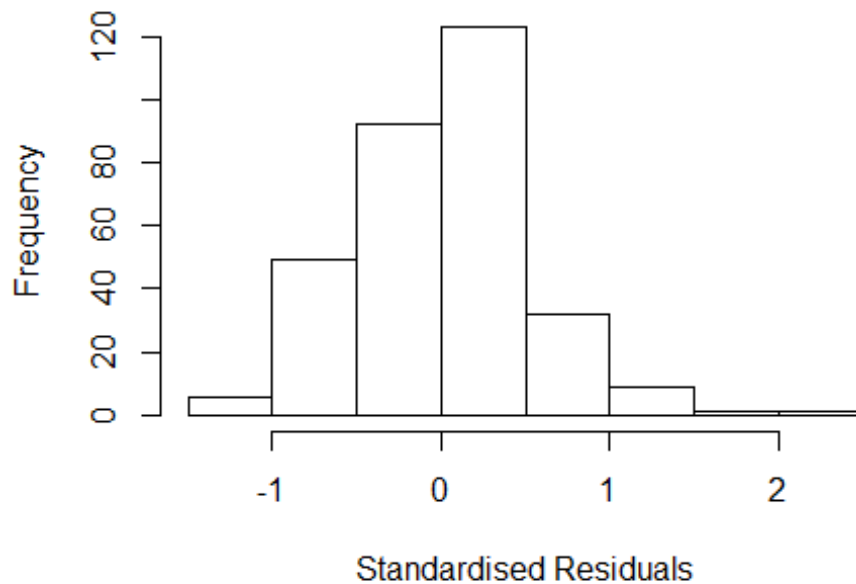
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 330, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.73674130620283"
## [1] "Male last author team size 2018 geometric mean: 3.44489770024044"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 240, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
##      Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.231 1          1.109
## LastAuthorFemale  1.528 1          1.236
## UniqueAuthors    3.412 4          1.166
## Year              4.605 16         1.049
```

## Residuals from first and last author and team size



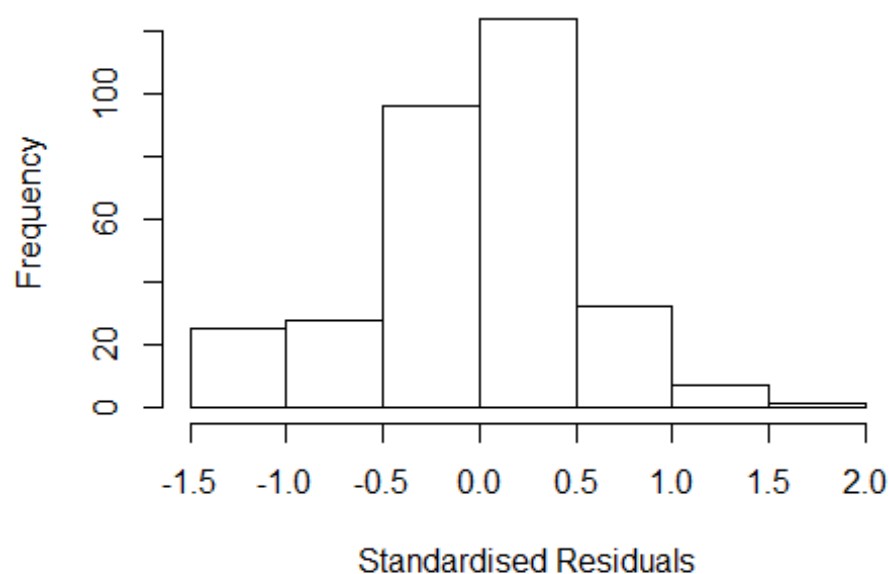
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4228 -0.3363 0.0306 0.3168 2.2979
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6569 0.2305 2.85 0.0047 **
## FirstAuthorFemale1 -0.0300 0.0592 -0.51 0.6130
## LastAuthorFemale1 -0.0961 0.0795 -1.21 0.2276
## UniqueAuthors2 0.4916 0.1500 3.28 0.0012 **
## UniqueAuthors3 0.6180 0.1461 4.23 3.1e-05 ***
## UniqueAuthors4 0.6324 0.1332 4.75 3.2e-06 ***
## UniqueAuthors5 0.6338 0.1291 4.91 1.5e-06 ***
## Year1997 0.3975 0.2572 1.55 0.1234
## Year1998 0.2680 0.2341 1.15 0.2531
## Year1999 0.1480 0.2159 0.69 0.4937
```

```

## Year2000          -0.1850      0.2385   -0.78   0.4385
## Year2001          -0.1600      0.2064   -0.78   0.4389
## Year2002           0.2475      0.1851    1.34   0.1824
## Year2003           0.1546      0.1968    0.79   0.4329
## Year2004           0.0402      0.1985    0.20   0.8397
## Year2005           0.2058      0.1999    1.03   0.3040
## Year2006          -0.0894      0.2063   -0.43   0.6651
## Year2007          -0.0836      0.1984   -0.42   0.6737
## Year2008           0.0257      0.2065    0.12   0.9010
## Year2009          -0.0546      0.2235   -0.24   0.8071
## Year2010           0.1583      0.2274    0.70   0.4869
## Year2011          -0.0841      0.2041   -0.41   0.6807
## Year2012          -0.0555      0.2221   -0.25   0.8029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.445
## Multiple R-squared:  0.282, Adjusted R-squared:  0.227
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## observation 248 is an outlier with |weight| = 0 ( < 0.00032);
## 29 weights are ~= 1. The remaining 283 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.104  0.857  0.943  0.888  0.982  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.19e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.170 1          1.082
## LastAuthorFemale  1.407 1          1.186
## Year              1.646 16          1.016

```

## Residuals from first and last author



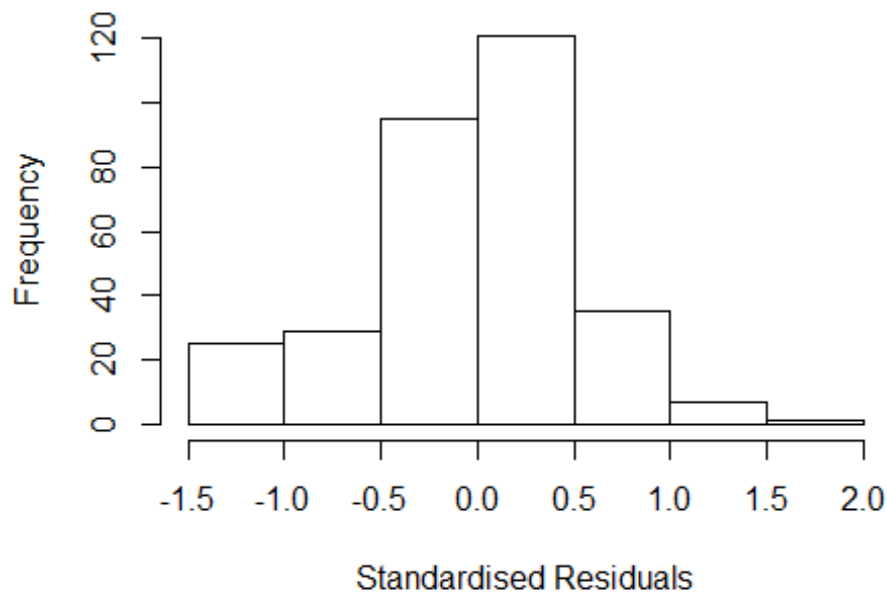
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3152 -0.3412 0.0489 0.3195 1.7029
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.16748 0.12842 9.09 <2e-16 ***
## FirstAuthorFemale1 0.04536 0.06386 0.71 0.478
## LastAuthorFemale1 -0.07647 0.08661 -0.88 0.378
## Year1997 0.45408 0.22946 1.98 0.049 *
## Year1998 0.32954 0.19515 1.69 0.092 .
## Year1999 0.13094 0.16635 0.79 0.432
## Year2000 -0.32172 0.20468 -1.57 0.117
## Year2001 -0.28732 0.23587 -1.22 0.224
## Year2002 0.23868 0.14064 1.70 0.091 .
## Year2003 0.16828 0.16158 1.04 0.299
## Year2004 0.02774 0.16803 0.17 0.869
## Year2005 0.19374 0.15162 1.28 0.202
```

```

## Year2006      -0.17024    0.15906   -1.07    0.285
## Year2007      -0.07179    0.15788   -0.45    0.650
## Year2008       0.00393    0.17287    0.02    0.982
## Year2009      -0.04627    0.22444   -0.21    0.837
## Year2010       0.14776    0.21119    0.70    0.485
## Year2011      -0.09128    0.17371   -0.53    0.600
## Year2012      -0.08926    0.18253   -0.49    0.625
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.485
## Multiple R-squared:  0.107, Adjusted R-squared:  0.0518
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 292 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.191  0.866  0.951  0.890  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.189 1      1.090
## Year              1.189 16      1.005

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3527 -0.3323 0.0615 0.3104 1.6343
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1699 0.1284 9.11 <2e-16 ***
## FirstAuthorFemale1 0.0401 0.0648 0.62 0.536
## Year1997 0.4409 0.2290 1.93 0.055 .
## Year1998 0.3231 0.1932 1.67 0.096 .
## Year1999 0.1061 0.1608 0.66 0.510
## Year2000 -0.3500 0.2015 -1.74 0.084 .
## Year2001 -0.2972 0.2404 -1.24 0.217
## Year2002 0.2277 0.1399 1.63 0.105
## Year2003 0.1683 0.1618 1.04 0.299
## Year2004 0.0240 0.1671 0.14 0.886
## Year2005 0.1600 0.1487 1.08 0.283
## Year2006 -0.1996 0.1534 -1.30 0.194
```

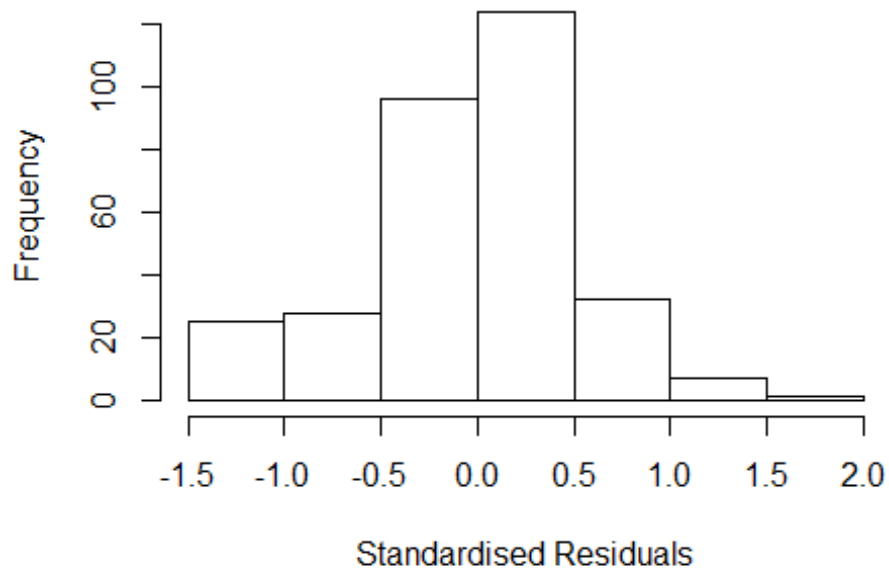


```

## Year2007          -0.0908      0.1552   -0.58    0.559
## Year2008          -0.0101      0.1720   -0.06    0.953
## Year2009          -0.0724      0.2236   -0.32    0.746
## Year2010           0.1427      0.2155    0.66    0.508
## Year2011          -0.1097      0.1727   -0.64    0.526
## Year2012          -0.1018      0.1844   -0.55    0.581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.484
## Multiple R-squared:  0.104, Adjusted R-squared:  0.0524
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 298 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.230  0.869  0.955  0.891  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.386 1      1.177
## Year              1.386 16      1.010

```

## Residuals from last author



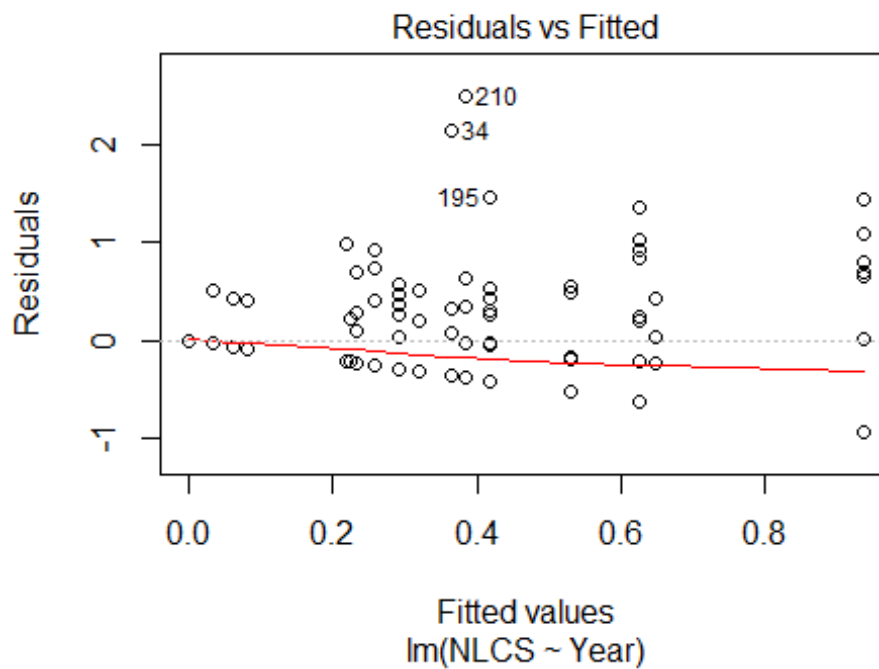
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3311 -0.3558 0.0447 0.3317 1.7275
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18827 0.12186 9.75 <2e-16 ***
## LastAuthorFemale1 -0.07159 0.08564 -0.84 0.404
## Year1997 0.46296 0.22994 2.01 0.045 *
## Year1998 0.32044 0.19338 1.66 0.099 .
## Year1999 0.12170 0.16342 0.74 0.457
## Year2000 -0.32043 0.20113 -1.59 0.112
## Year2001 -0.29029 0.23686 -1.23 0.221
## Year2002 0.23204 0.13819 1.68 0.094 .
## Year2003 0.16832 0.16400 1.03 0.306
## Year2004 0.02880 0.16858 0.17 0.864
## Year2005 0.19102 0.14928 1.28 0.202
## Year2006 -0.16815 0.15647 -1.07 0.283
```

```

## Year2007          -0.08037      0.15657    -0.51      0.608
## Year2008           0.00192      0.17284      0.01      0.991
## Year2009          -0.04986      0.22262    -0.22      0.823
## Year2010           0.14280      0.21346      0.67      0.504
## Year2011          -0.09611      0.17311    -0.56      0.579
## Year2012          -0.09317      0.18078    -0.52      0.607
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.483
## Multiple R-squared:  0.105, Adjusted R-squared:  0.0537
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 294 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.173  0.868  0.951  0.889  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.19e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 313"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3001"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9   19   4   11   16   12   12   9   11   19   15   4   8   15   13
## 2011 2012
##   21   14
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    8   12   4   10   16   11   9   8   11   18   14   4   6   14   13
## 2011 2012

```

```
## 14 11
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 8 11 4 9 16 11 9 7 11 18 12 4 5 14 13
## 2011 2012
## 13 10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = Inf, df = 16, p-value <2e-16
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.3, df = 1, p-value = 0.02
## [1] "Female first author team size 2018 geometric mean: 5.47722557505166"
## [1] "Male first author team size 2018 geometric mean: 3.14519819078556"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 11, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.95789213552895"
## [1] "Male last author team size 2018 geometric mean: 2.47166234431194"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

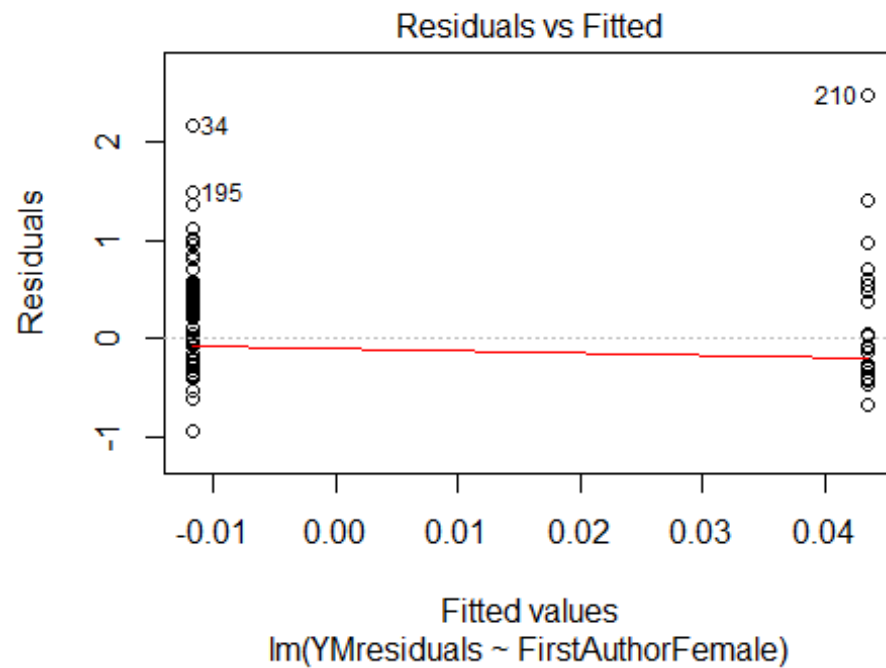
## Warning in lmrob.S(x, y, control = control, mf = mf): find_scale() did not
## converge in 'maxit.scale' (= 200) iterations

## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"

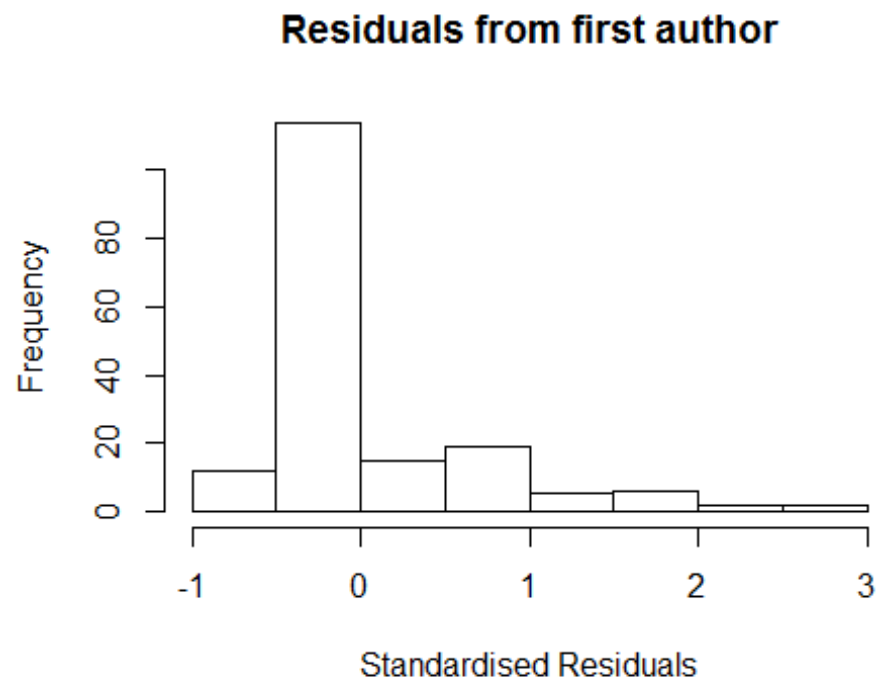
## Warning in lmrob.S(x, y, control = control, mf = mf): S-estimated scale ==
## 0: Probably exact fit; check your data

## Warning in cov2cor(v): diag(.) had 0 or NA entries; non-finite result is
## doubtful

```



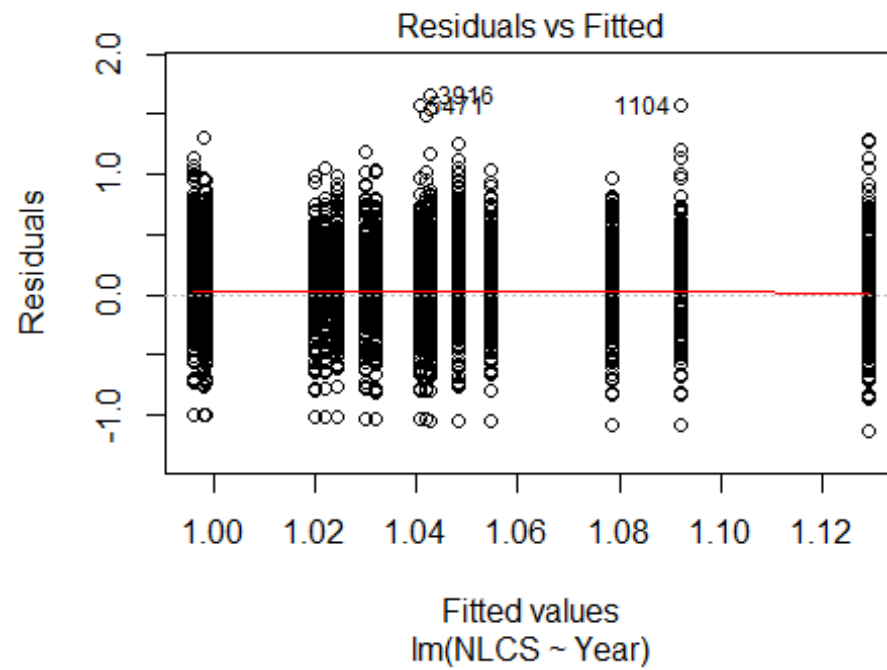
##		GVIF	Df	GVIF^(1/(2*Df))
##	FirstAuthorFemale	NaN	1	NaN
##	Year	NaN	16	NaN



```

## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 175"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3002"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 511 488 495 486 498 449 423 390 441 363 422 376 339 398 417
## 2011 2012
## 419 365
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 380 360 380 366 318 295 330 287 332 280 343 300 288 331 327
## 2011 2012
## 340 285
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 335 316 337 330 285 254 303 256 302 257 299 274 252 305 297
## 2011 2012
## 318 251
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 100, df = 16, p-value = 8e-15

```

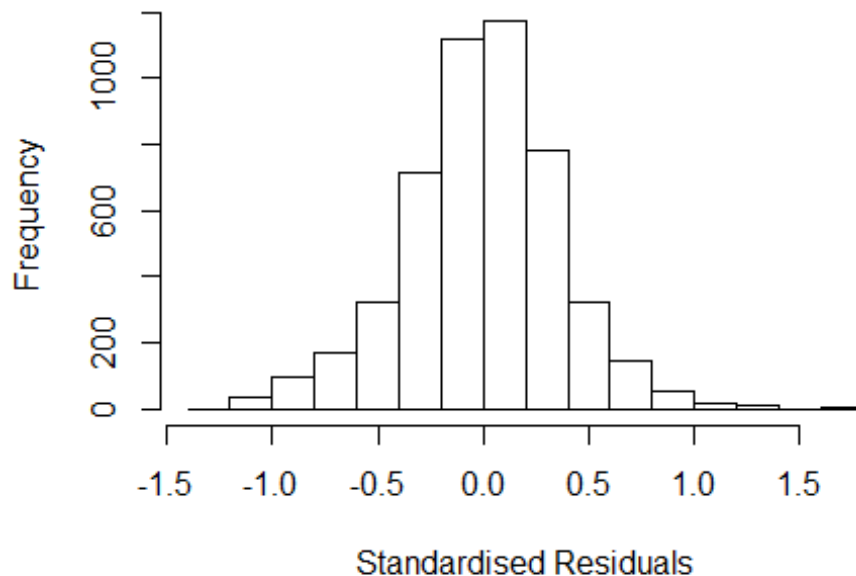


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.8, df = 1, p-value = 0.2
```





## Residuals from first and last author and team size



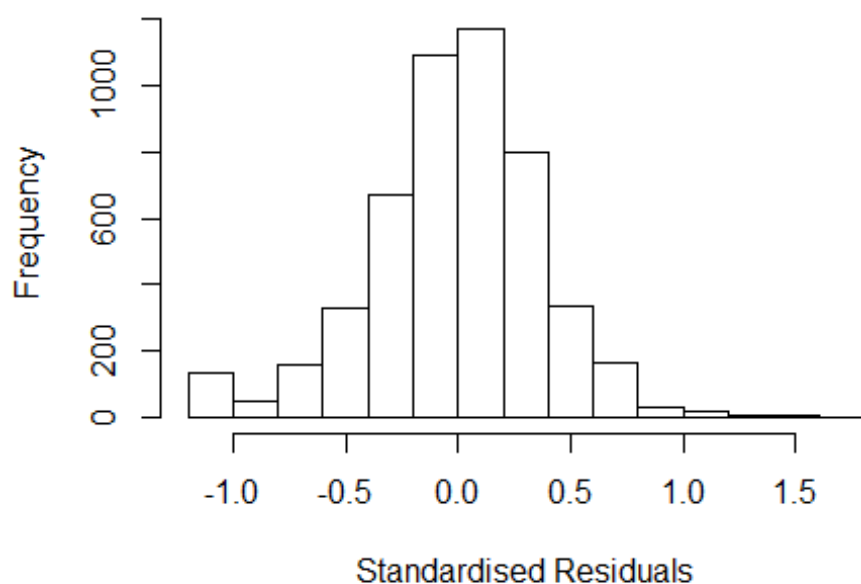
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22197 -0.21794  0.00342  0.22148  1.73286
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.91382    0.04347   21.02 < 2e-16 ***
## FirstAuthorFemale1 -0.00499    0.01260   -0.40  0.69176
## LastAuthorFemale1 -0.00820    0.01513   -0.54  0.58769
## UniqueAuthors2     0.15447    0.04038    3.83  0.00013 ***
## UniqueAuthors3     0.18451    0.03966    4.65  3.4e-06 ***
## UniqueAuthors4     0.21578    0.03997    5.40  7.0e-08 ***
## UniqueAuthors5     0.31859    0.03908    8.15  4.5e-16 ***
## Year1997          -0.01043    0.03116   -0.33  0.73779
## Year1998          -0.05155    0.02870   -1.80  0.07258 .
## Year1999          -0.03809    0.02822   -1.35  0.17719
```

```

## Year2000      -0.07411      0.02853      -2.60      0.00940 **
## Year2001      -0.10492      0.02972      -3.53      0.00042 ***
## Year2002      -0.05968      0.02885      -2.07      0.03861 *
## Year2003      -0.11241      0.03038      -3.70      0.00022 ***
## Year2004      -0.08134      0.02849      -2.86      0.00432 **
## Year2005      -0.15266      0.03010      -5.07      4.1e-07 ***
## Year2006      -0.08351      0.02933      -2.85      0.00442 **
## Year2007      -0.11705      0.02947      -3.97      7.2e-05 ***
## Year2008      -0.12404      0.03146      -3.94      8.2e-05 ***
## Year2009      -0.12509      0.03118      -4.01      6.1e-05 ***
## Year2010      -0.14314      0.02995      -4.78      1.8e-06 ***
## Year2011      -0.12847      0.03464      -3.71      0.00021 ***
## Year2012      -0.10181      0.03743      -2.72      0.00655 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.33
## Multiple R-squared:  0.0733, Adjusted R-squared:  0.0692
## Convergence in 21 IRWLS iterations
##
## Robustness weights:
## 4 observations c(2141,2143,2995,3314)
## are outliers with |weight| = 0 ( < 2e-05);
## 386 weights are ~ = 1. The remaining 4581 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0145 0.8660 0.9530 0.8900 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.040 1      1.020
## LastAuthorFemale 1.028 1      1.014
## Year      1.058 16      1.002

```

## Residuals from first and last author



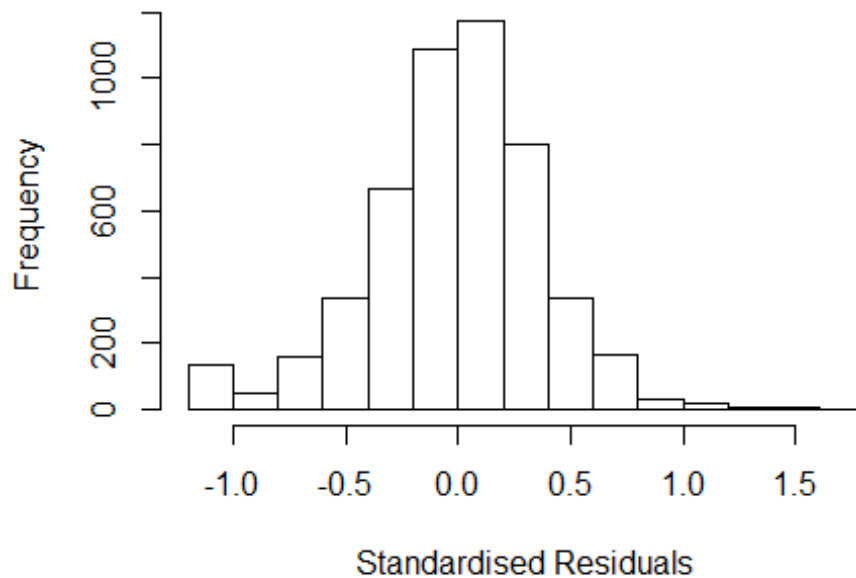
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13115 -0.22509 0.00631 0.22080 1.62791
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.128073 0.021942 51.41 < 2e-16 ***
## FirstAuthorFemale1 0.003077 0.012808 0.24 0.81013
## LastAuthorFemale1 0.000559 0.015416 0.04 0.97106
## Year1997 -0.018508 0.030974 -0.60 0.55019
## Year1998 -0.057818 0.028812 -2.01 0.04483 *
## Year1999 -0.039930 0.028455 -1.40 0.16060
## Year2000 -0.082057 0.028187 -2.91 0.00362 **
## Year2001 -0.096405 0.030276 -3.18 0.00146 **
## Year2002 -0.057984 0.028330 -2.05 0.04074 *
## Year2003 -0.092730 0.030688 -3.02 0.00253 **
## Year2004 -0.069540 0.028598 -2.43 0.01507 *
## Year2005 -0.131045 0.030899 -4.24 2.3e-05 ***
```

```

## Year2006          -0.054785    0.029261    -1.87    0.06122 .
## Year2007          -0.092687    0.030802    -3.01    0.00263 **
## Year2008          -0.086418    0.031697    -2.73    0.00643 **
## Year2009          -0.088384    0.031243    -2.83    0.00469 **
## Year2010          -0.117110    0.030202    -3.88    0.00011 ***
## Year2011          -0.103847    0.033755    -3.08    0.00211 **
## Year2012          -0.073010    0.039157    -1.86    0.06231 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.335
## Multiple R-squared:  0.00892,    Adjusted R-squared:  0.00531
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2143 is an outlier with |weight| = 0 ( < 2e-05);
## 414 weights are ~= 1. The remaining 4556 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  0.862  0.951   0.886   0.985   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           2.01e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.035 1         1.017
## Year              1.035 16         1.001

```

## Residuals from first author



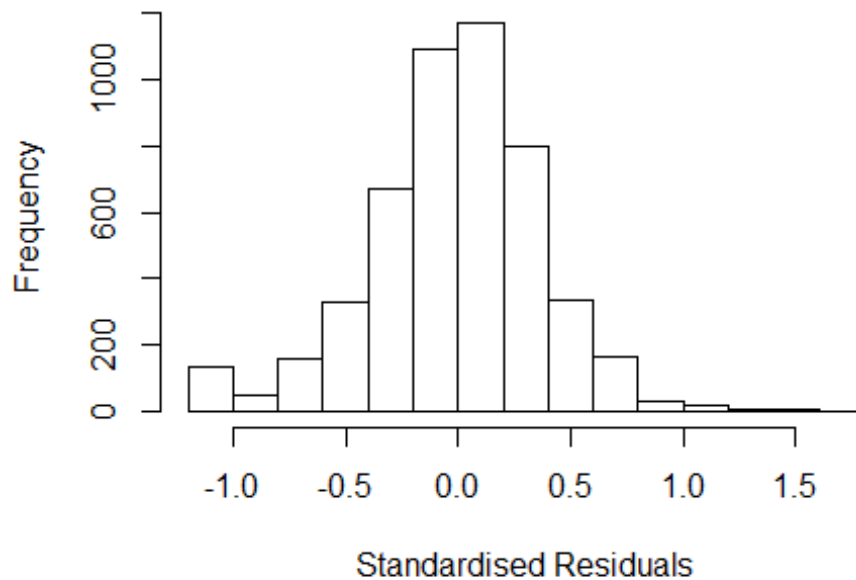
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.13122 -0.22519 0.00619 0.22083 1.62781
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12815 0.02182 51.70 < 2e-16 ***
## FirstAuthorFemale1 0.00308 0.01280 0.24 0.81020
## Year1997 -0.01851 0.03095 -0.60 0.54980
## Year1998 -0.05781 0.02881 -2.01 0.04486 *
## Year1999 -0.03991 0.02846 -1.40 0.16091
## Year2000 -0.08207 0.02819 -2.91 0.00361 **
## Year2001 -0.09642 0.03024 -3.19 0.00144 **
## Year2002 -0.05795 0.02833 -2.05 0.04084 *
## Year2003 -0.09274 0.03067 -3.02 0.00251 **
## Year2004 -0.06954 0.02860 -2.43 0.01506 *
## Year2005 -0.13103 0.03090 -4.24 2.3e-05 ***
## Year2006 -0.05477 0.02926 -1.87 0.06131 .
```

```

## Year2007          -0.09269    0.03080   -3.01  0.00263 **
## Year2008          -0.08639    0.03168   -2.73  0.00642 **
## Year2009          -0.08833    0.03122   -2.83  0.00469 **
## Year2010          -0.11709    0.03021   -3.88  0.00011 ***
## Year2011          -0.10381    0.03374   -3.08  0.00210 **
## Year2012          -0.07298    0.03916   -1.86  0.06247 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.335
## Multiple R-squared:  0.00892,    Adjusted R-squared:  0.00551
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2143 is an outlier with |weight| = 0 ( < 2e-05);
## 414 weights are ~= 1. The remaining 4556 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  0.862  0.951  0.886  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.022 1          1.011
## Year          1.022 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.12856 -0.22570 0.00661 0.22131 1.62730
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.128556 0.021849 51.65 < 2e-16 ***
## LastAuthorFemale1 0.000545 0.015416 0.04 0.97178
## Year1997 -0.018652 0.030957 -0.60 0.54687
## Year1998 -0.057862 0.028814 -2.01 0.04468 *
## Year1999 -0.039866 0.028450 -1.40 0.16120
## Year2000 -0.081877 0.028160 -2.91 0.00366 **
## Year2001 -0.096416 0.030272 -3.18 0.00146 **
## Year2002 -0.057852 0.028320 -2.04 0.04112 *
## Year2003 -0.092521 0.030649 -3.02 0.00255 **
## Year2004 -0.069363 0.028593 -2.43 0.01531 *
## Year2005 -0.130794 0.030843 -4.24 2.3e-05 ***
## Year2006 -0.054610 0.029238 -1.87 0.06186 .
```

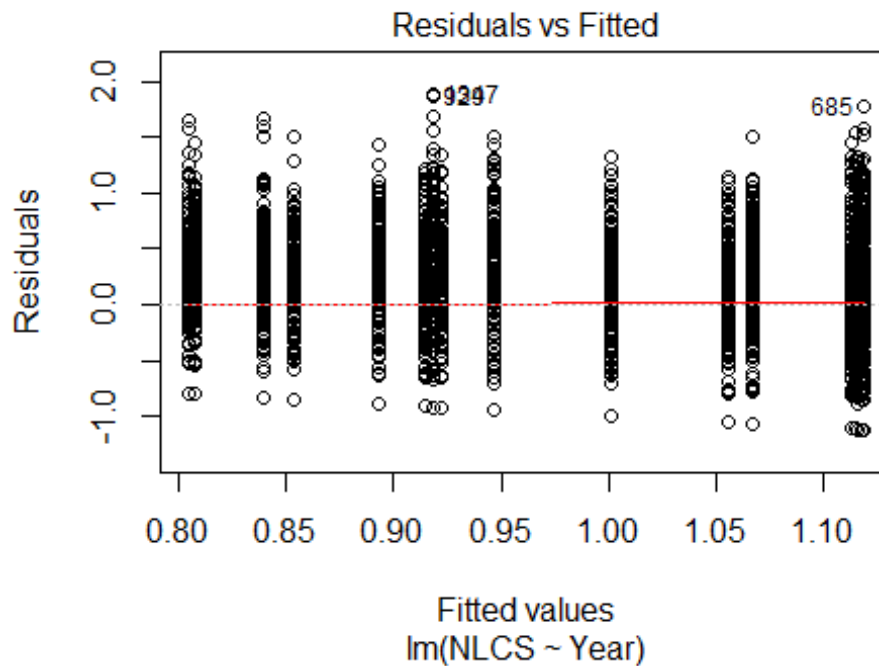


```

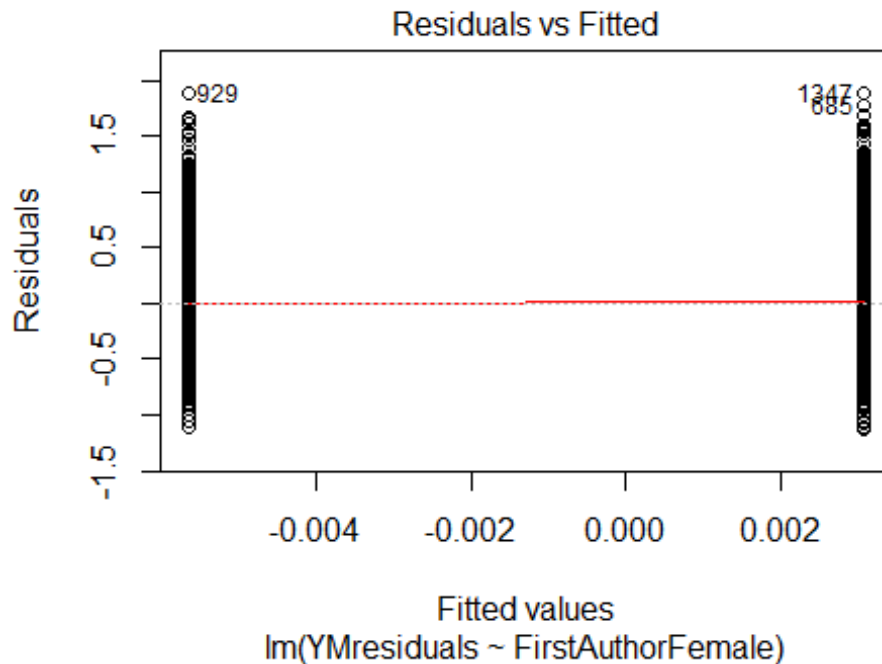
## Year2007          -0.092376    0.030758    -3.00    0.00268 **
## Year2008          -0.086163    0.031652    -2.72    0.00651 **
## Year2009          -0.088265    0.031245    -2.82    0.00475 **
## Year2010          -0.116589    0.030101    -3.87    0.00011 ***
## Year2011          -0.103663    0.033746    -3.07    0.00214 **
## Year2012          -0.072598    0.039089    -1.86    0.06334 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.335
## Multiple R-squared:  0.0089, Adjusted R-squared:  0.0055
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 2143 is an outlier with |weight| = 0 ( < 2e-05);
## 415 weights are ~= 1. The remaining 4555 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0001 0.8630 0.9510 0.8860 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.01e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4971"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3003"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 422 369 493 481 512 278 298 296 335 334 365 350 446 321 322
## 2011 2012
## 397 358
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 199 169 211 166 208 109 198 211 241 253 275 273 331 246 221

```

```
## 2011 2012
## 295 262
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 174 151 194 147 180 102 178 189 218 235 254 245 301 230 191
## 2011 2012
## 255 223
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 70, df = 16, p-value = 1e-08
```

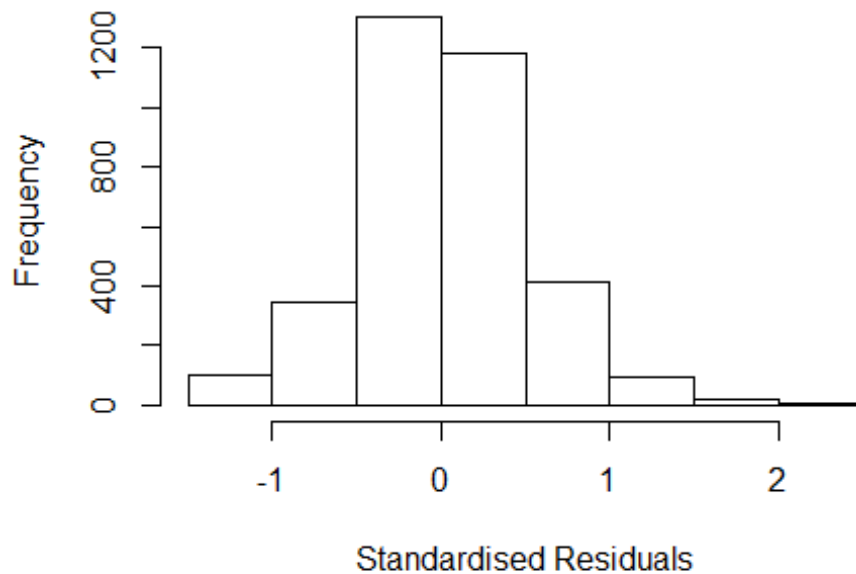


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
```



```
## [1] "Female first author team size 2018 geometric mean: 3.91902561142126"
## [1] "Male first author team size 2018 geometric mean: 3.8364974624239"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5500, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.73450247137849"
## [1] "Male last author team size 2018 geometric mean: 3.92153717644546"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4300, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.166 1 1.080
## LastAuthorFemale 1.143 1 1.069
## UniqueAuthors 1.364 4 1.040
## Year 1.350 16 1.009
```

## Residuals from first and last author and team size



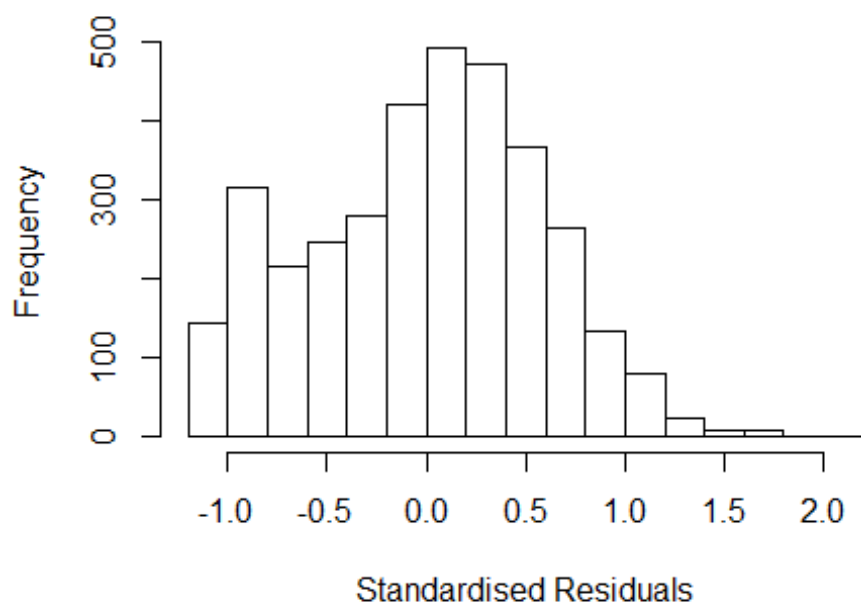
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.33735 -0.28052 -0.00713 0.32437 2.33458
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.46568 0.05291 8.80 < 2e-16 ***
## FirstAuthorFemale1 0.00973 0.01926 0.51 0.61340
## LastAuthorFemale1 -0.01842 0.02020 -0.91 0.36189
## UniqueAuthors2 0.48828 0.03673 13.29 < 2e-16 ***
## UniqueAuthors3 0.79391 0.03280 24.20 < 2e-16 ***
## UniqueAuthors4 0.85377 0.03312 25.78 < 2e-16 ***
## UniqueAuthors5 0.85291 0.02357 36.18 < 2e-16 ***
## Year1997 0.06802 0.08027 0.85 0.39681
## Year1998 -0.18926 0.07050 -2.68 0.00729 **
## Year1999 -0.03164 0.06982 -0.45 0.65042
```

```

## Year2000      -0.16669    0.07479   -2.23  0.02588 *
## Year2001      -0.01795    0.06974   -0.26  0.79688
## Year2002      -0.21693    0.05751   -3.77  0.00016 ***
## Year2003      -0.16061    0.05783   -2.78  0.00551 **
## Year2004      -0.14191    0.05858   -2.42  0.01546 *
## Year2005      -0.22187    0.05648   -3.93  8.7e-05 ***
## Year2006      -0.20554    0.05660   -3.63  0.00029 ***
## Year2007      -0.18516    0.05741   -3.23  0.00127 **
## Year2008      -0.25725    0.05435   -4.73  2.3e-06 ***
## Year2009      -0.17965    0.05947   -3.02  0.00254 **
## Year2010      -0.15097    0.05772   -2.62  0.00895 **
## Year2011      -0.10332    0.05779   -1.79  0.07389 .
## Year2012      -0.11113    0.05784   -1.92  0.05478 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.435
## Multiple R-squared:  0.351, Adjusted R-squared:  0.346
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 3 observations c(469,471,1013)
## are outliers with |weight| <= 1.3e-06 ( < 2.9e-05);
## 276 weights are ~= 1. The remaining 3188 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0003 0.8440 0.9500 0.8810 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1 1.022
## LastAuthorFemale 1.047 1 1.023
## Year 1.056 16 1.002

```

## Residuals from first and last author



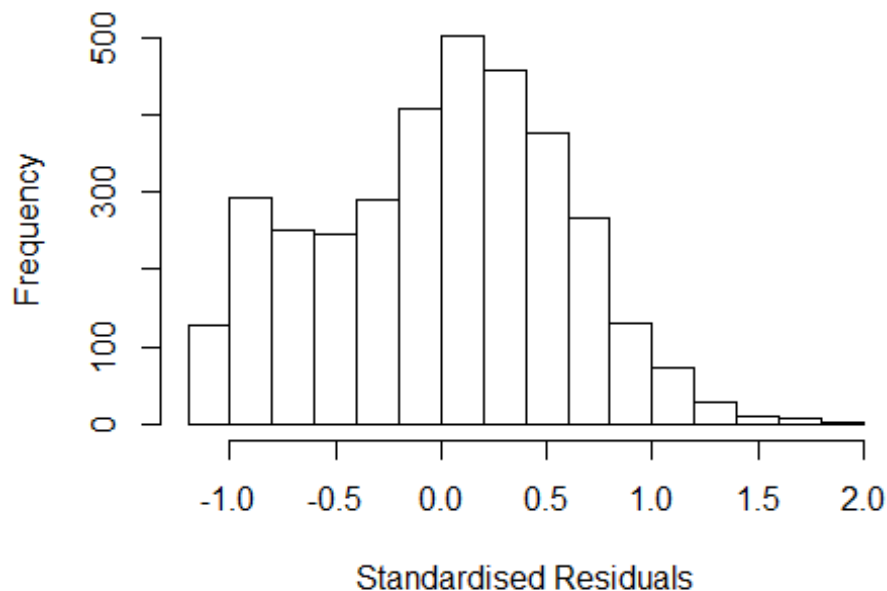
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1559 -0.4359 0.0411 0.4071 2.0101
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.107573 0.053464 20.72 < 2e-16 ***
## FirstAuthorFemale1 0.021258 0.021703 0.98 0.3274
## LastAuthorFemale1 -0.113476 0.024802 -4.58 4.9e-06 ***
## Year1997 0.023290 0.085288 0.27 0.7848
## Year1998 -0.231464 0.078266 -2.96 0.0031 **
## Year1999 0.007804 0.075881 0.10 0.9181
## Year2000 -0.119375 0.076136 -1.57 0.1170
## Year2001 0.027083 0.074035 0.37 0.7145
## Year2002 -0.235039 0.068400 -3.44 0.0006 ***
## Year2003 -0.160501 0.064351 -2.49 0.0127 *
## Year2004 -0.194208 0.067314 -2.89 0.0039 **
## Year2005 -0.312412 0.068062 -4.59 4.6e-06 ***
```

```

## Year2006      -0.263198    0.064471   -4.08  4.6e-05 ***
## Year2007      -0.173219    0.067175   -2.58  0.0100 **
## Year2008      -0.300433    0.064349   -4.67  3.1e-06 ***
## Year2009      -0.099708    0.064089   -1.56  0.1199
## Year2010      -0.031982    0.063389   -0.50  0.6139
## Year2011      -0.000773    0.063300   -0.01  0.9903
## Year2012       0.003405    0.062875    0.05  0.9568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.591
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0405
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 281 weights are ~= 1. The remaining 3186 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.223  0.854  0.946  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1308 -0.4412 0.0475 0.4115 1.9361
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.094552 0.052790 20.73 < 2e-16 ***
## FirstAuthorFemale1 -0.002917 0.021950 -0.13 0.89427
## Year1997 0.022433 0.084565 0.27 0.79081
## Year1998 -0.227690 0.077642 -2.93 0.00338 **
## Year1999 0.002958 0.075434 0.04 0.96872
## Year2000 -0.118562 0.075677 -1.57 0.11728
## Year2001 0.036298 0.073532 0.49 0.62159
## Year2002 -0.248535 0.067963 -3.66 0.00026 ***
## Year2003 -0.168091 0.064245 -2.62 0.00893 **
## Year2004 -0.203973 0.067521 -3.02 0.00254 **
## Year2005 -0.318422 0.068024 -4.68 3.0e-06 ***
## Year2006 -0.269462 0.064121 -4.20 2.7e-05 ***
```

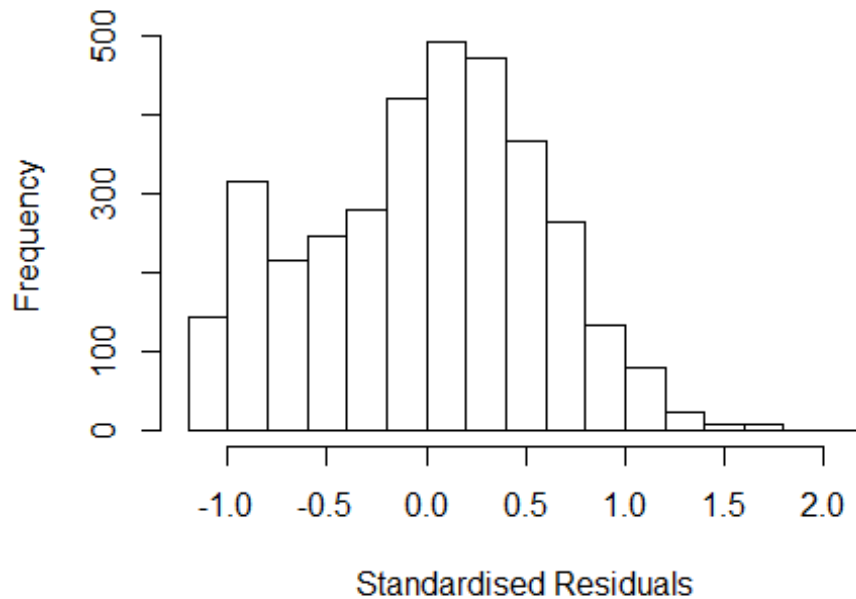


```

## Year2007          -0.187807    0.066894    -2.81    0.00502 **
## Year2008          -0.315102    0.064510    -4.88    1.1e-06 ***
## Year2009          -0.100634    0.063706    -1.58    0.11428
## Year2010          -0.035357    0.062808    -0.56    0.57351
## Year2011          -0.002126    0.062833    -0.03    0.97301
## Year2012           0.000407    0.062531     0.01    0.99481
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.593
## Multiple R-squared:  0.0397, Adjusted R-squared:  0.035
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 286 weights are ~= 1. The remaining 3181 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.264  0.850  0.944  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.032 1          1.016
## Year            1.032 16          1.001

```

## Residuals from last author



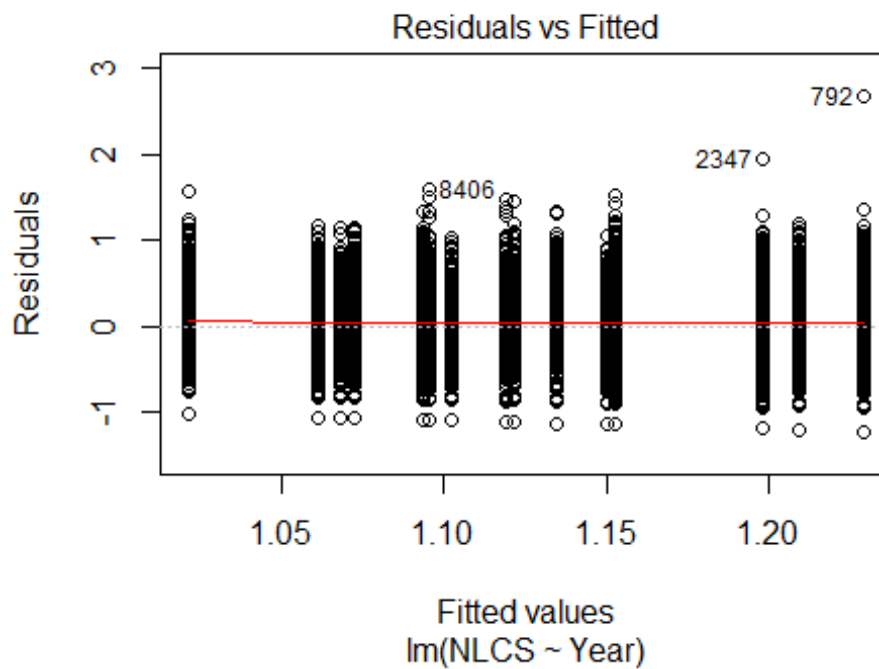
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1411 -0.4355  0.0409  0.4081  2.0192
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11396    0.05304   21.00 < 2e-16 ***
## LastAuthorFemale1 -0.10752    0.02469   -4.35 1.4e-05 ***
## Year1997         0.02290    0.08539    0.27 0.78854
## Year1998        -0.23164    0.07823   -2.96 0.00309 **
## Year1999         0.00662    0.07599    0.09 0.93063
## Year2000        -0.11867    0.07616   -1.56 0.11927
## Year2001         0.02710    0.07415    0.37 0.71476
## Year2002        -0.23448    0.06855   -3.42 0.00063 ***
## Year2003        -0.16120    0.06439   -2.50 0.01235 *
## Year2004        -0.19406    0.06747   -2.88 0.00405 **
## Year2005        -0.31239    0.06820   -4.58 4.8e-06 ***
## Year2006        -0.26260    0.06453   -4.07 4.8e-05 ***
```

```

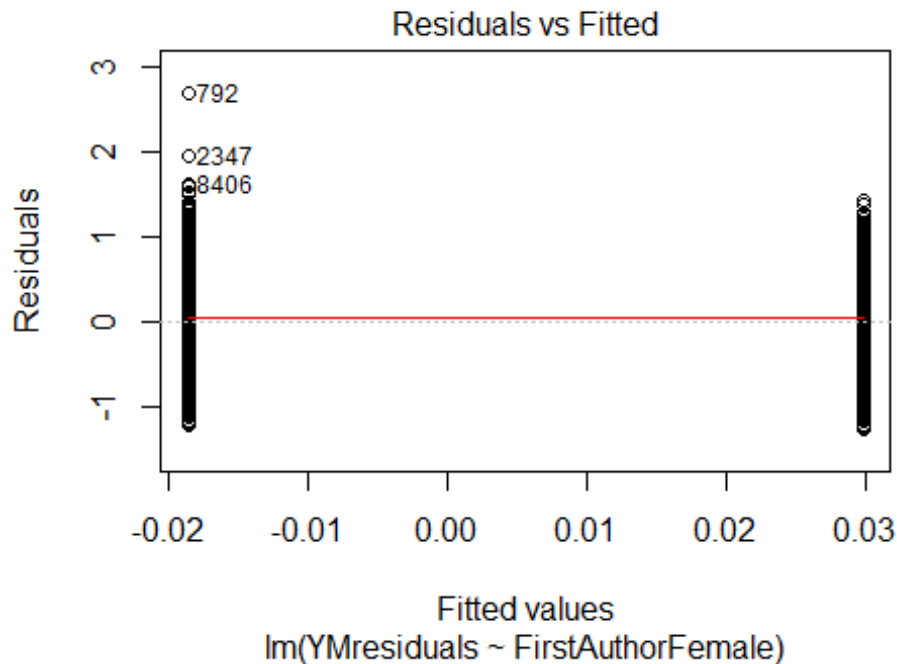
## Year2007          -0.17239      0.06730      -2.56   0.01046 *
## Year2008          -0.30082      0.06447      -4.67   3.2e-06 ***
## Year2009          -0.10024      0.06423      -1.56   0.11871
## Year2010          -0.03275      0.06349      -0.52   0.60605
## Year2011          -0.00218      0.06343      -0.03   0.97252
## Year2012           0.00371      0.06305       0.06   0.95313
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.592
## Multiple R-squared:  0.0452, Adjusted R-squared:  0.0405
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 284 weights are ~= 1. The remaining 3183 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.220  0.853  0.945  0.910  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.88e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3467"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3004"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1129 1102 1149 1004 1114 892 799 683 699 614 681 619 700 586 601
## 2011 2012
## 650 666
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 518 491 522 422 449 309 480 433 474 443 449 383 468 402 381
## 2011 2012

```

```
## 472 473
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 463 445 470 375 403 280 431 380 418 388 389 329 416 341 332
## 2011 2012
## 419 415
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 96, df = 16, p-value = 2e-13
```

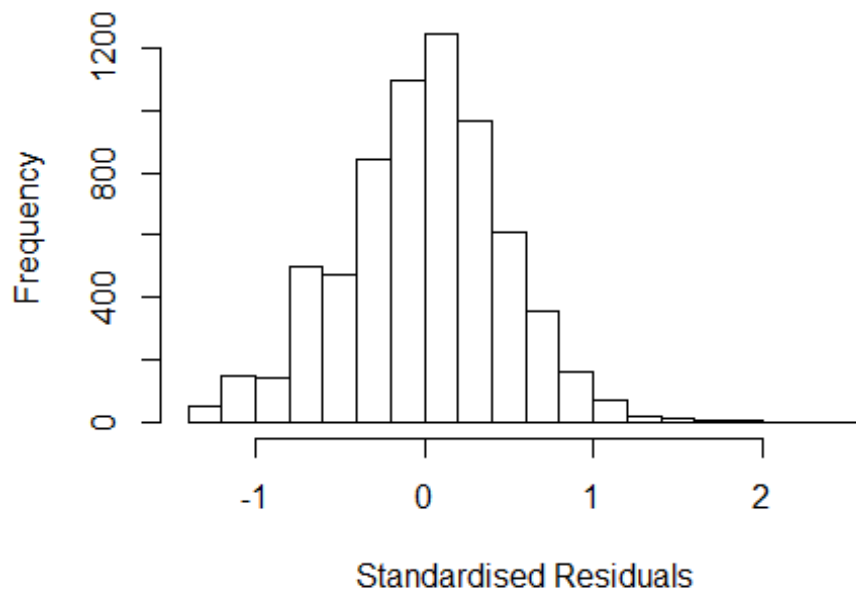


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 25, df = 1, p-value = 5e-07
```



```
## [1] "Female first author team size 2018 geometric mean: 4.72639347378856"
## [1] "Male first author team size 2018 geometric mean: 4.19193820867171"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 10000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.83140959852407"
## [1] "Male last author team size 2018 geometric mean: 4.3010673048014"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8000, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1      1.022
## LastAuthorFemale  1.029 1      1.014
## UniqueAuthors    1.130 4      1.015
## Year              1.121 16     1.004
```

## Residuals from first and last author and team size



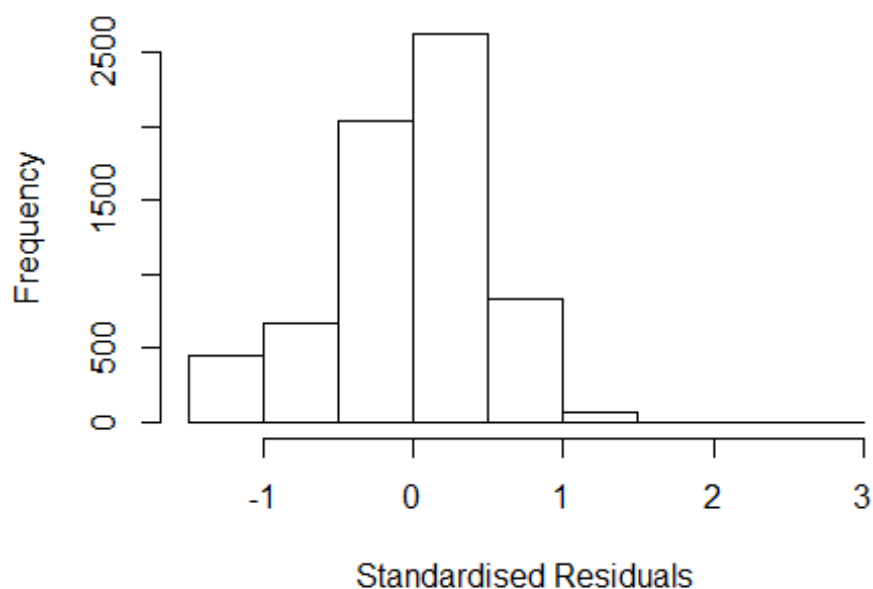
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 0029914622 3.907 1996      3004      1      2.535
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3768 -0.3051  0.0164  0.3053  2.5346
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.92e-01   3.33e-02  23.75  < 2e-16 ***
## FirstAuthorFemale1  9.72e-05   1.21e-02   0.01  0.99360
## LastAuthorFemale1 -2.31e-03   1.52e-02  -0.15  0.87928
## UniqueAuthors2     3.84e-01   2.96e-02  12.96  < 2e-16 ***
## UniqueAuthors3     4.99e-01   2.84e-02  17.57  < 2e-16 ***
## UniqueAuthors4     5.34e-01   2.84e-02  18.78  < 2e-16 ***
## UniqueAuthors5     5.81e-01   2.67e-02  21.73  < 2e-16 ***
## Year1997           4.47e-03   3.61e-02   0.12  0.90161
## Year1998          -8.49e-02   3.50e-02  -2.43  0.01528 *
## Year1999          -6.51e-02   3.56e-02  -1.83  0.06790 .
```

```

## Year2000      -1.26e-01  3.53e-02  -3.58  0.00035 ***
## Year2001      -9.18e-02  3.52e-02  -2.61  0.00916 **
## Year2002      -1.22e-01  3.33e-02  -3.66  0.00026 ***
## Year2003      -1.54e-01  3.29e-02  -4.69  2.8e-06 ***
## Year2004      -1.42e-01  3.18e-02  -4.48  7.6e-06 ***
## Year2005      -1.52e-01  3.32e-02  -4.57  4.9e-06 ***
## Year2006      -1.48e-01  3.22e-02  -4.61  4.1e-06 ***
## Year2007      -1.18e-01  3.39e-02  -3.47  0.00053 ***
## Year2008      -1.85e-01  3.70e-02  -4.99  6.3e-07 ***
## Year2009      -1.31e-01  3.46e-02  -3.80  0.00015 ***
## Year2010      -1.30e-01  3.45e-02  -3.78  0.00016 ***
## Year2011      -2.02e-02  3.37e-02  -0.60  0.54791
## Year2012      -9.24e-02  3.41e-02  -2.71  0.00672 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.444
## Multiple R-squared:  0.151, Adjusted R-squared:  0.148
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 2 observations c(307,825) are outliers with |weight| = 0 ( < 1.5e-05);
## 577 weights are ~ = 1. The remaining 6115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0219 0.8480 0.9480 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev mts compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1 1.008
## LastAuthorFemale 1.019 1 1.009
## Year 1.031 16 1.001

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 0029914622 3.907 1996      3004      1      2.682
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2754 -0.3121  0.0307  0.3179  2.6823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22473    0.02622   46.72 < 2e-16 ***
## FirstAuthorFemale1 0.05066    0.01262    4.01 6.0e-05 ***
## LastAuthorFemale1 -0.01083    0.01628   -0.67 0.50605
## Year1997        -0.01964    0.03863   -0.51 0.61114
## Year1998        -0.10079    0.03691   -2.73 0.00633 **
## Year1999        -0.05780    0.03780   -1.53 0.12628
## Year2000        -0.12713    0.03683   -3.45 0.00056 ***
## Year2001        -0.07261    0.03627   -2.00 0.04530 *
## Year2002        -0.13203    0.03578   -3.69 0.00023 ***
## Year2003        -0.15657    0.03596   -4.35 1.4e-05 ***
## Year2004        -0.13459    0.03362   -4.00 6.3e-05 ***
## Year2005        -0.13894    0.03610   -3.85 0.00012 ***
```

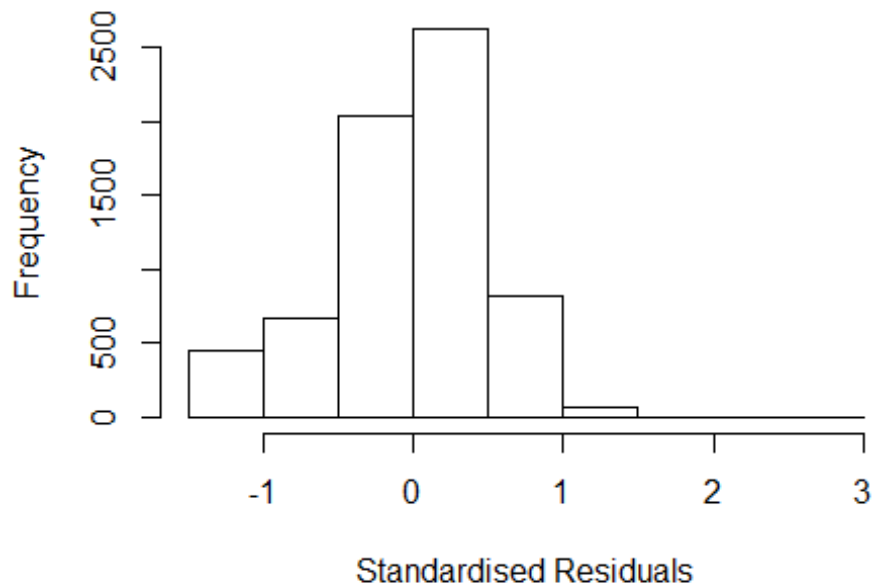


```

## Year2006          -0.14860      0.03476      -4.27  1.9e-05 ***
## Year2007          -0.11711      0.03555      -3.29  0.00099 ***
## Year2008          -0.20233      0.04131      -4.90  9.9e-07 ***
## Year2009          -0.08970      0.03608      -2.49  0.01295 *
## Year2010          -0.10859      0.03664      -2.96  0.00305 **
## Year2011          -0.00243      0.03429      -0.07  0.94350
## Year2012          -0.08107      0.03621      -2.24  0.02518 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.0121
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 307 is an outlier with |weight| = 0 ( < 1.5e-05);
## 540 weights are ~= 1. The remaining 6153 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0368 0.8530 0.9490 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.015 1          1.007
## Year              1.015 16          1.000

```

## Residuals from first author



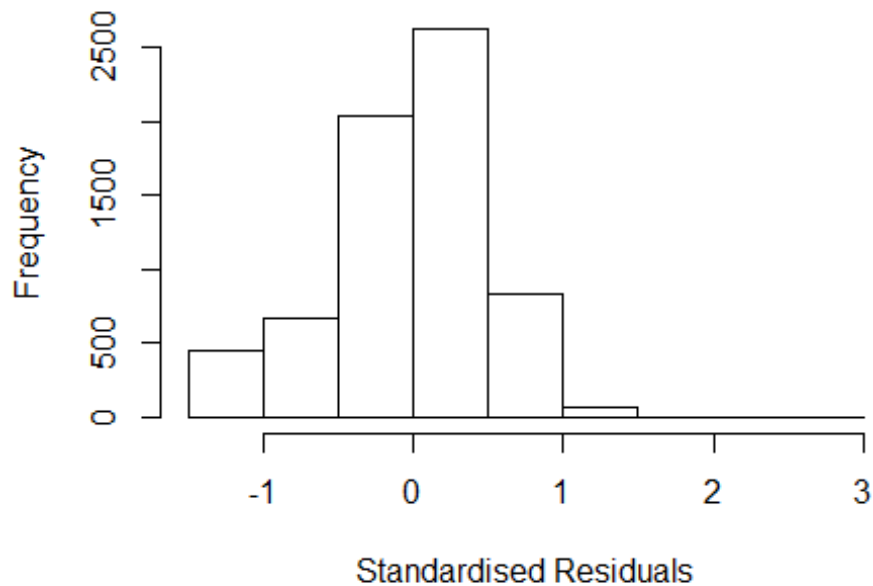
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 0029914622 3.907 1996      3004      1      2.682
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2738 -0.3113  0.0309  0.3179  2.6832
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.22383    0.02616   46.78 < 2e-16 ***
## FirstAuthorFemale1 0.04992    0.01272    3.92 8.8e-05 ***
## Year1997        -0.02039    0.03864   -0.53 0.59764
## Year1998        -0.10122    0.03689   -2.74 0.00609 **
## Year1999        -0.05883    0.03774   -1.56 0.11907
## Year2000        -0.12844    0.03673   -3.50 0.00047 ***
## Year2001        -0.07326    0.03625   -2.02 0.04333 *
## Year2002        -0.13291    0.03576   -3.72 0.00020 ***
## Year2003        -0.15747    0.03595   -4.38 1.2e-05 ***
## Year2004        -0.13553    0.03359   -4.03 5.5e-05 ***
## Year2005        -0.14023    0.03600   -3.90 9.9e-05 ***
## Year2006        -0.14974    0.03472   -4.31 1.6e-05 ***
```

```

## Year2007          -0.11882    0.03543   -3.35  0.00080 ***
## Year2008          -0.20377    0.04134   -4.93  8.4e-07 ***
## Year2009          -0.09058    0.03603   -2.51  0.01196 *
## Year2010          -0.10926    0.03662   -2.98  0.00286 **
## Year2011          -0.00355    0.03422   -0.10  0.91727
## Year2012          -0.08217    0.03618   -2.27  0.02316 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.0123
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 307 is an outlier with |weight| = 0 ( < 1.5e-05);
## 542 weights are ~= 1. The remaining 6151 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0363 0.8530 0.9490 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.016 1          1.008
## Year          1.016 16          1.001

```

## Residuals from last author



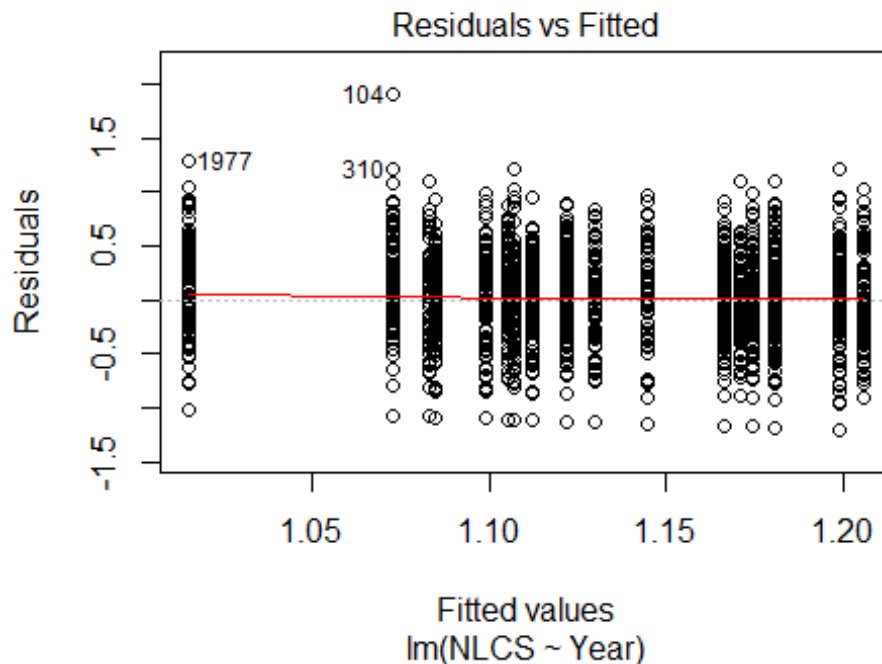
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 792 0029914622 3.907 1996      3004      1      2.682
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2443 -0.3141  0.0326  0.3178  2.6627
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.24431    0.02548   48.83  < 2e-16 ***
## LastAuthorFemale1 -0.00555    0.01634   -0.34  0.73406
## Year1997          -0.02277    0.03868   -0.59  0.55615
## Year1998          -0.10065    0.03690   -2.73  0.00640 **
## Year1999          -0.06182    0.03775   -1.64  0.10154
## Year2000          -0.12958    0.03680   -3.52  0.00043 ***
## Year2001          -0.07165    0.03626   -1.98  0.04823 *
## Year2002          -0.13343    0.03577   -3.73  0.00019 ***
## Year2003          -0.15683    0.03592   -4.37  1.3e-05 ***
## Year2004          -0.13539    0.03359   -4.03  5.6e-05 ***
## Year2005          -0.13890    0.03609   -3.85  0.00012 ***
## Year2006          -0.14719    0.03469   -4.24  2.2e-05 ***
```

```

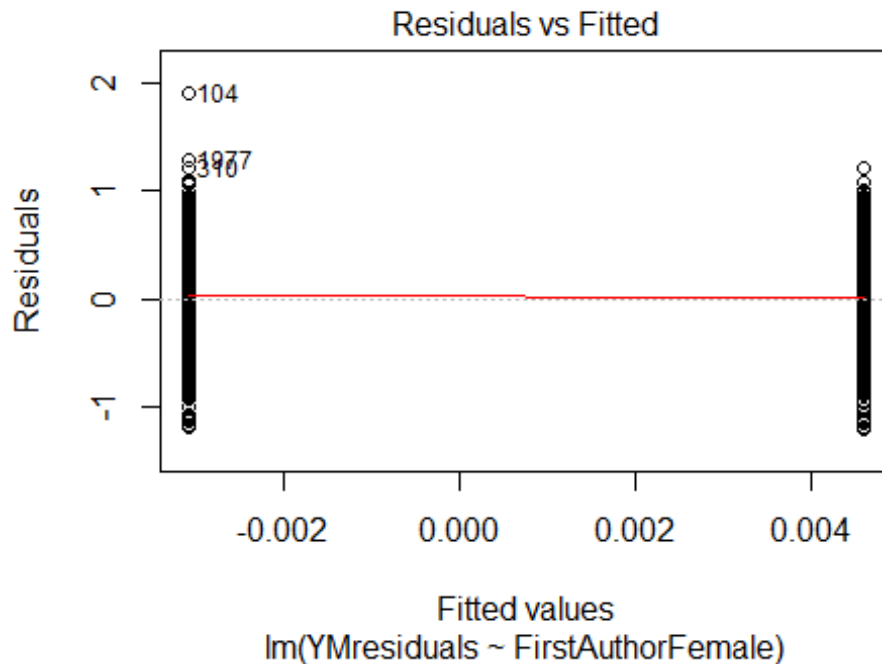
## Year2007          -0.11685      0.03562    -3.28   0.00104 **
## Year2008          -0.20304      0.04120    -4.93   8.5e-07 ***
## Year2009          -0.08660      0.03610    -2.40   0.01647 *
## Year2010          -0.10843      0.03683    -2.94   0.00325 **
## Year2011          -0.00480      0.03412    -0.14   0.88804
## Year2012          -0.08013      0.03625    -2.21   0.02711 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.461
## Multiple R-squared:  0.0124, Adjusted R-squared:  0.00987
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 307 is an outlier with |weight| = 0 ( < 1.5e-05);
## 549 weights are ~= 1. The remaining 6144 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0423 0.8520 0.9490 0.8890 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6694"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3005"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 315 301 294 236 215 249 252 208 185 190 198 216 240 213 206
## 2011 2012
## 197 191
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 132 142 120 92 68 73 124 110 88 122 110 126 151 139 138

```

```
## 2011 2012
## 124 130
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 122 134 106 82 57 68 114 99 72 100 100 111 133 119 127
## 2011 2012
## 115 106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 90, df = 16, p-value = 2e-12
```

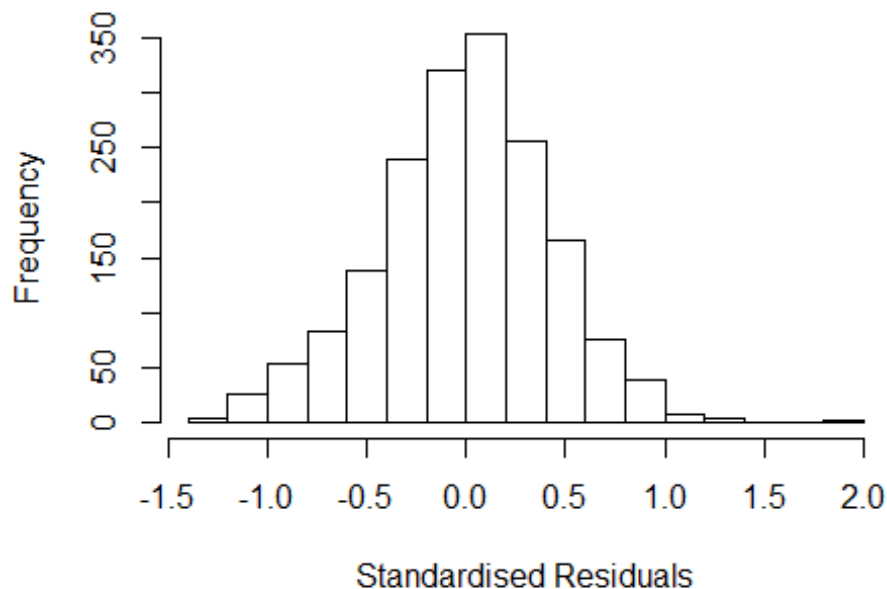


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.3, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 5.06654026410202"
## [1] "Male first author team size 2018 geometric mean: 4.70550695111769"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1500, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.66764727207159"
## [1] "Male last author team size 2018 geometric mean: 4.94455698346747"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.084 1          1.041
## LastAuthorFemale  1.044 1          1.022
## UniqueAuthors     1.208 4          1.024
## Year               1.268 16         1.007
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22634 -0.27859  0.00993  0.27228  1.90031
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.99487    0.07164   13.89  < 2e-16 ***
## FirstAuthorFemale1 0.00228    0.02137    0.11   0.915
## LastAuthorFemale1 -0.01587    0.02372   -0.67   0.504
## UniqueAuthors2    0.07882    0.05117    1.54   0.124
## UniqueAuthors3    0.11653    0.04988    2.34   0.020 *
## UniqueAuthors4    0.13574    0.05314    2.55   0.011 *
## UniqueAuthors5    0.19387    0.04860    3.99 6.9e-05 ***
## Year1997          0.09345    0.06916    1.35   0.177
## Year1998          0.10046    0.07150    1.40   0.160
## Year1999          0.09385    0.07399    1.27   0.205
```

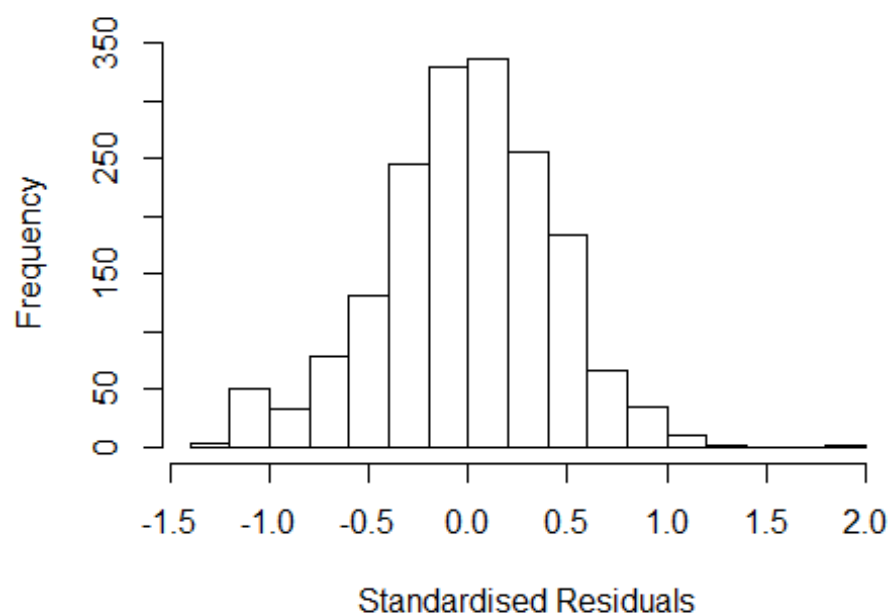


```

## Year2000      0.08202      0.08682      0.94      0.345
## Year2001     -0.02414      0.08123     -0.30      0.766
## Year2002     -0.08460      0.07541     -1.12      0.262
## Year2003     -0.02561      0.06587     -0.39      0.698
## Year2004      0.00308      0.07352      0.04      0.967
## Year2005      0.01137      0.06992      0.16      0.871
## Year2006      0.05172      0.06473      0.80      0.424
## Year2007     -0.02550      0.06474     -0.39      0.694
## Year2008      0.03404      0.06037      0.56      0.573
## Year2009     -0.01089      0.06290     -0.17      0.863
## Year2010      0.00868      0.06501      0.13      0.894
## Year2011      0.05571      0.06445      0.86      0.388
## Year2012      0.00391      0.06970      0.06      0.955
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.412
## Multiple R-squared:  0.0311, Adjusted R-squared:  0.0188
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 152 weights are ~= 1. The remaining 1613 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0009 0.8620 0.9510 0.8980 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.078 1      1.038
## LastAuthorFemale 1.021 1      1.010
## Year      1.094 16      1.003

```

## Residuals from first and last author



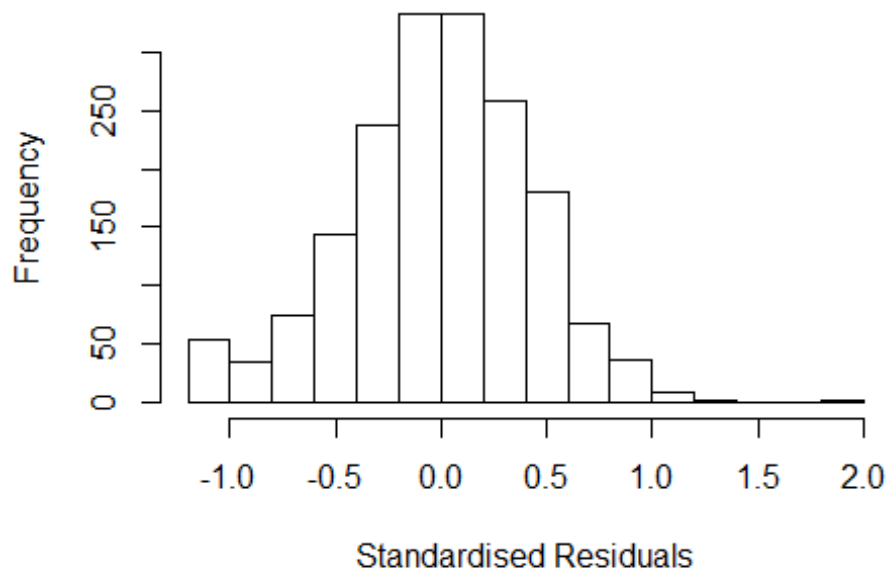
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20731 -0.28455 0.00482 0.27739 1.87162
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10238 0.05727 19.25 <2e-16 ***
## FirstAuthorFemale1 0.00713 0.02143 0.33 0.74
## LastAuthorFemale1 -0.02457 0.02374 -1.03 0.30
## Year1997 0.09780 0.07039 1.39 0.16
## Year1998 0.10761 0.07240 1.49 0.14
## Year1999 0.09579 0.07512 1.28 0.20
## Year2000 0.08895 0.08959 0.99 0.32
## Year2001 -0.00407 0.08219 -0.05 0.96
## Year2002 -0.06836 0.07777 -0.88 0.38
## Year2003 -0.00514 0.06808 -0.08 0.94
## Year2004 0.02221 0.07556 0.29 0.77
## Year2005 0.03345 0.07195 0.46 0.64
```

```

## Year2006          0.06529      0.06641      0.98      0.33
## Year2007         -0.00435      0.06634     -0.07      0.95
## Year2008          0.06023      0.06252      0.96      0.34
## Year2009          0.01717      0.06506      0.26      0.79
## Year2010          0.02286      0.06714      0.34      0.73
## Year2011          0.07371      0.06632      1.11      0.27
## Year2012          0.02714      0.07177      0.38      0.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00179
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 166 weights are ~= 1. The remaining 1599 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0042 0.8600 0.9490 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.074 1      1.036
## Year              1.074 16      1.002

```

## Residuals from first author



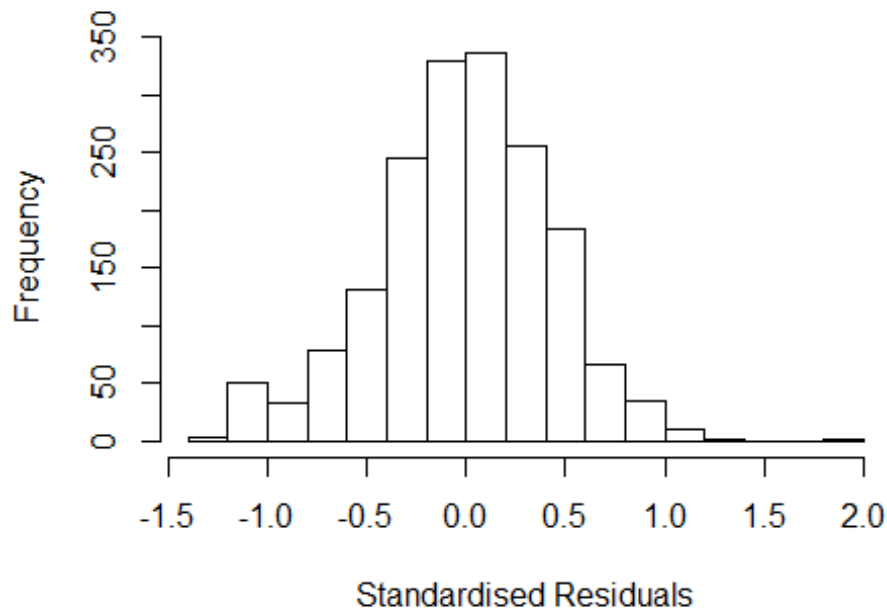
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.19915 -0.28439 0.00624 0.27475 1.87550
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09850 0.05702 19.27 <2e-16 ***
## FirstAuthorFemale1 0.00497 0.02141 0.23 0.82
## Year1997 0.09569 0.07032 1.36 0.17
## Year1998 0.10556 0.07229 1.46 0.14
## Year1999 0.09489 0.07520 1.26 0.21
## Year2000 0.08722 0.08948 0.97 0.33
## Year2001 -0.00532 0.08221 -0.06 0.95
## Year2002 -0.06984 0.07772 -0.90 0.37
## Year2003 -0.00776 0.06819 -0.11 0.91
## Year2004 0.02152 0.07574 0.28 0.78
## Year2005 0.03190 0.07203 0.44 0.66
## Year2006 0.06511 0.06638 0.98 0.33
```

```

## Year2007      -0.00564    0.06629   -0.09    0.93
## Year2008      0.05759    0.06252    0.92    0.36
## Year2009      0.01674    0.06501    0.26    0.80
## Year2010      0.02130    0.06710    0.32    0.75
## Year2011      0.07330    0.06621    1.11    0.27
## Year2012      0.02475    0.07206    0.34    0.73
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00178
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 165 weights are ~= 1. The remaining 1600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0036 0.8580 0.9480 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1          1.009
## Year              1.017 16          1.001

```

## Residuals from last author



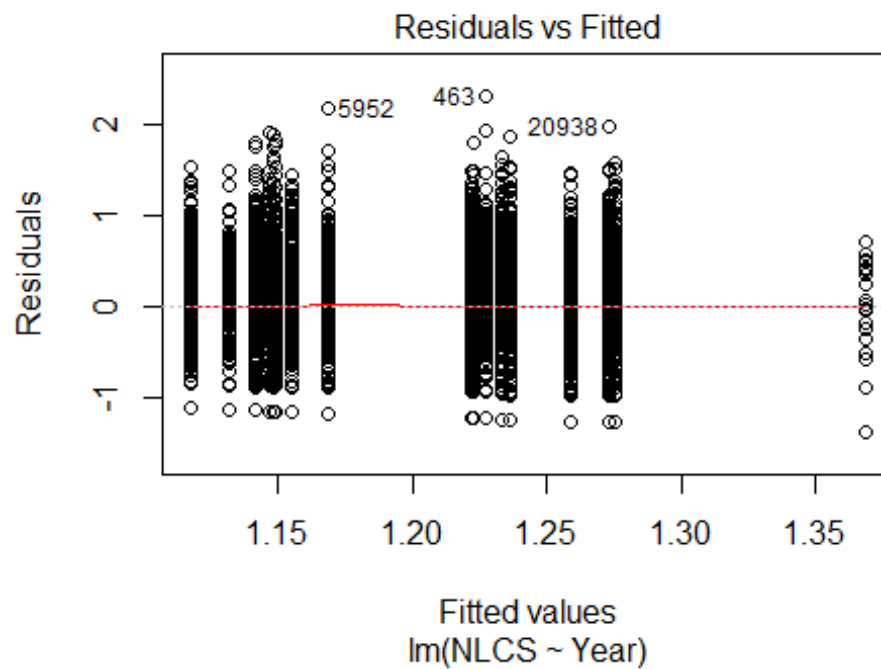
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.20224 -0.28404 0.00474 0.27603 1.86892
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.10508 0.05618 19.67 <2e-16 ***
## LastAuthorFemale1 -0.02378 0.02371 -1.00 0.32
## Year1997 0.09716 0.07031 1.38 0.17
## Year1998 0.10740 0.07235 1.48 0.14
## Year1999 0.09617 0.07529 1.28 0.20
## Year2000 0.08744 0.08897 0.98 0.33
## Year2001 -0.00445 0.08227 -0.05 0.96
## Year2002 -0.06819 0.07778 -0.88 0.38
## Year2003 -0.00557 0.06804 -0.08 0.93
## Year2004 0.02204 0.07557 0.29 0.77
## Year2005 0.03351 0.07199 0.47 0.64
## Year2006 0.06567 0.06653 0.99 0.32
```

```

## Year2007          -0.00411      0.06641    -0.06      0.95
## Year2008           0.06053      0.06257      0.97      0.33
## Year2009           0.01796      0.06515      0.28      0.78
## Year2010           0.02295      0.06721      0.34      0.73
## Year2011           0.07438      0.06643      1.12      0.26
## Year2012           0.02765      0.07184      0.38      0.70
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0119, Adjusted R-squared:  0.00224
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 165 weights are ~= 1. The remaining 1600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0044 0.8590 0.9490 0.8960 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1765"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3100"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   99  937  969 1118  851 1084 1034 1075  896  885  973  973  972 1001  856
## 2011 2012
##   893  857
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   22  228  277  369  271  267  381  419  348  367  398  434  452  490  424
## 2011 2012

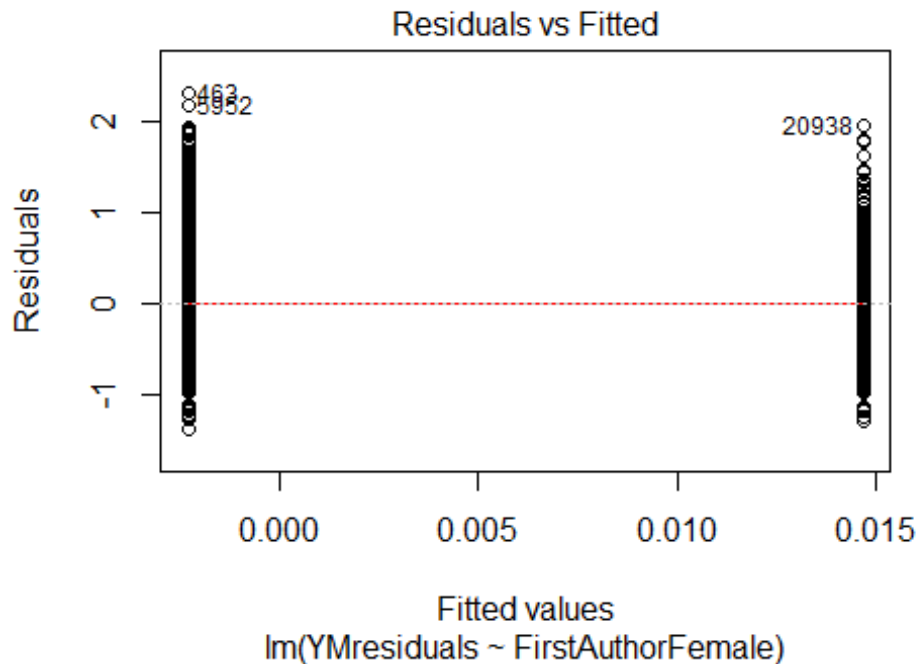
```

```
## 428 444
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 20 213 256 333 248 242 344 384 320 330 357 389 388 448 368
## 2011 2012
## 379 391
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```



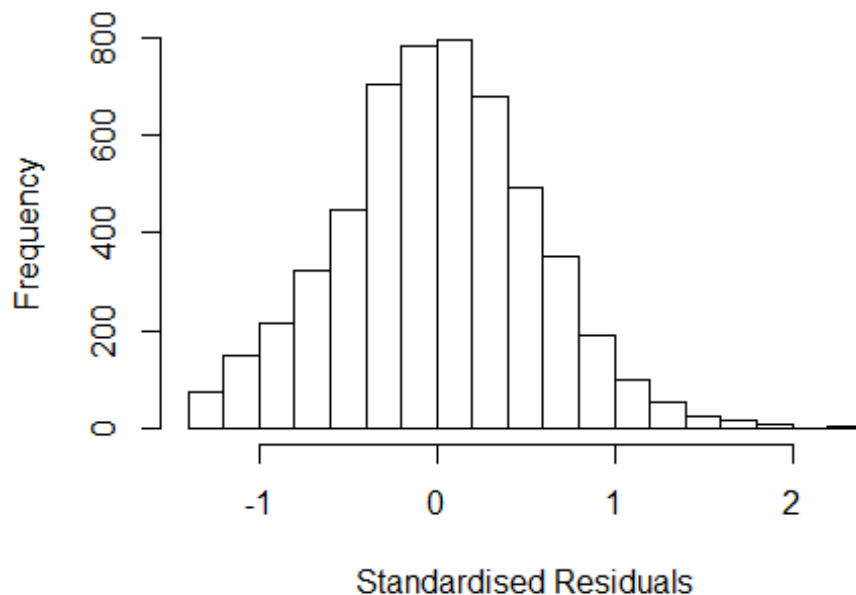
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.2, df = 1, p-value = 0.04
```





```
## [1] "Female first author team size 2018 geometric mean: 4.47894548215175"
## [1] "Male first author team size 2018 geometric mean: 2.78858447184675"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 30000, p-value = 5e-09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.963259034139"
## [1] "Male last author team size 2018 geometric mean: 2.94023230193137"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23000, p-value = 3e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.065 1          1.032
## LastAuthorFemale  1.057 1          1.028
## UniqueAuthors    1.123 4          1.015
## Year             1.130 16          1.004
```

## Residuals from first and last author and team size



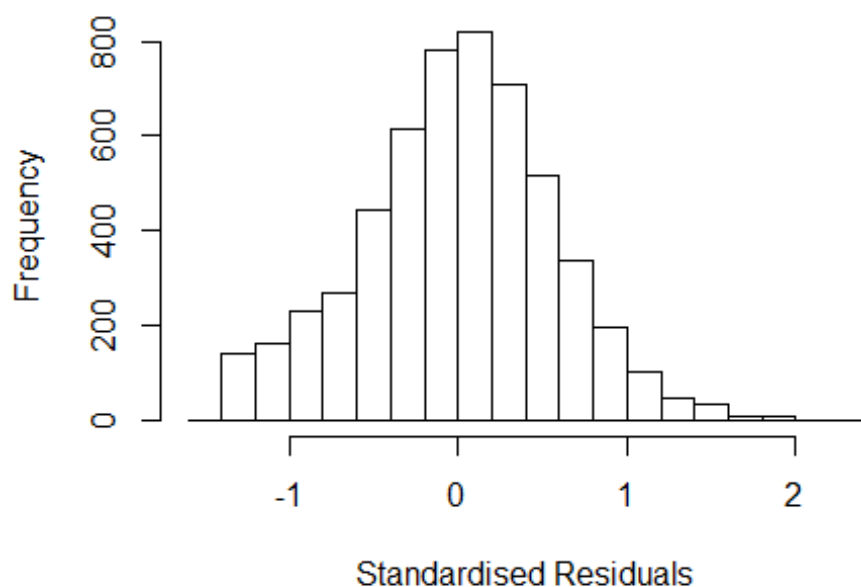
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3944 -0.3559 0.0037 0.3573 2.3104
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2822 0.1140 11.25 <2e-16 ***
## FirstAuthorFemale1 -0.0175 0.0212 -0.83 0.4093
## LastAuthorFemale1 0.0189 0.0273 0.69 0.4890
## UniqueAuthors2 0.2214 0.0210 10.52 <2e-16 ***
## UniqueAuthors3 0.2308 0.0239 9.67 <2e-16 ***
## UniqueAuthors4 0.3090 0.0278 11.12 <2e-16 ***
## UniqueAuthors5 0.3459 0.0266 13.02 <2e-16 ***
## Year1997 -0.2472 0.1199 -2.06 0.0392 *
## Year1998 -0.1874 0.1188 -1.58 0.1148
## Year1999 -0.2336 0.1169 -2.00 0.0458 *
```

```

## Year2000          -0.3240      0.1176   -2.75   0.0059 **
## Year2001          -0.2506      0.1207   -2.08   0.0380 *
## Year2002          -0.2259      0.1168   -1.93   0.0532 .
## Year2003          -0.2984      0.1178   -2.53   0.0113 *
## Year2004          -0.3194      0.1178   -2.71   0.0067 **
## Year2005          -0.3457      0.1172   -2.95   0.0032 **
## Year2006          -0.3338      0.1172   -2.85   0.0044 **
## Year2007          -0.3122      0.1169   -2.67   0.0076 **
## Year2008          -0.3075      0.1166   -2.64   0.0084 **
## Year2009          -0.2384      0.1162   -2.05   0.0402 *
## Year2010          -0.2439      0.1160   -2.10   0.0356 *
## Year2011          -0.2680      0.1163   -2.30   0.0213 *
## Year2012          -0.2515      0.1166   -2.16   0.0311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.0574, Adjusted R-squared:  0.0535
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 476 weights are ~= 1. The remaining 4934 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0222 0.8630 0.9510 0.9000 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.85e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.039 1 1.019
## LastAuthorFemale 1.045 1 1.022
## Year 1.029 16 1.001

```

## Residuals from first and last author



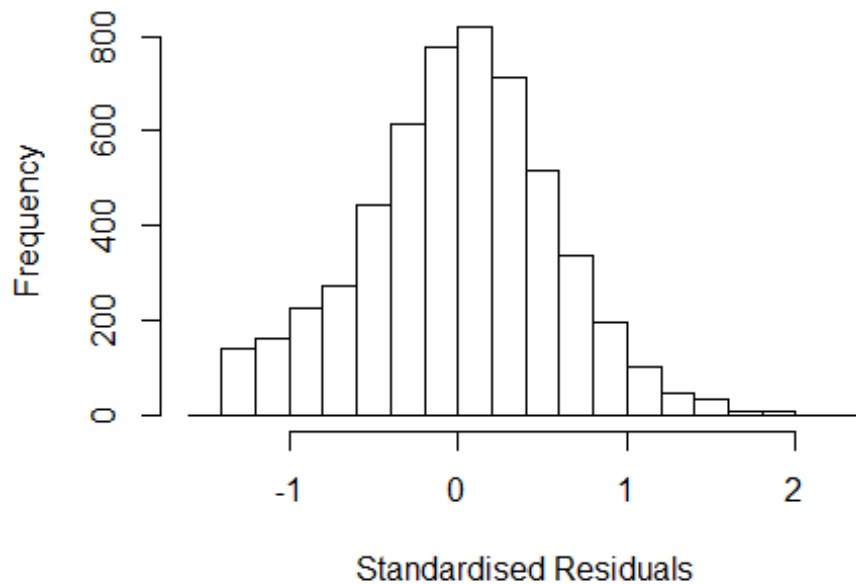
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4591 -0.3641 0.0153 0.3642 2.3419
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.45906 0.11271 12.94 <2e-16 ***
## FirstAuthorFemale1 0.02757 0.02133 1.29 0.1962
## LastAuthorFemale1 0.00883 0.02791 0.32 0.7517
## Year1997 -0.25900 0.11990 -2.16 0.0308 *
## Year1998 -0.20583 0.11837 -1.74 0.0821 .
## Year1999 -0.24244 0.11639 -2.08 0.0373 *
## Year2000 -0.33053 0.11713 -2.82 0.0048 **
## Year2001 -0.25872 0.12040 -2.15 0.0317 *
## Year2002 -0.22838 0.11622 -1.96 0.0495 *
## Year2003 -0.30880 0.11722 -2.63 0.0085 **
## Year2004 -0.32279 0.11724 -2.75 0.0059 **
## Year2005 -0.37026 0.11707 -3.16 0.0016 **
```

```

## Year2006      -0.34490    0.11708   -2.95   0.0032 **
## Year2007      -0.32951    0.11643   -2.83   0.0047 **
## Year2008      -0.33405    0.11637   -2.87   0.0041 **
## Year2009      -0.22770    0.11612   -1.96   0.0499 *
## Year2010      -0.21018    0.11587   -1.81   0.0697 .
## Year2011      -0.24425    0.11617   -2.10   0.0356 *
## Year2012      -0.21639    0.11647   -1.86   0.0632 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.0072
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 479 weights are ~= 1. The remaining 4931 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0206 0.8590 0.9500 0.8980 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.013 1      1.006
## Year      1.013 16      1.000

```

## Residuals from first author



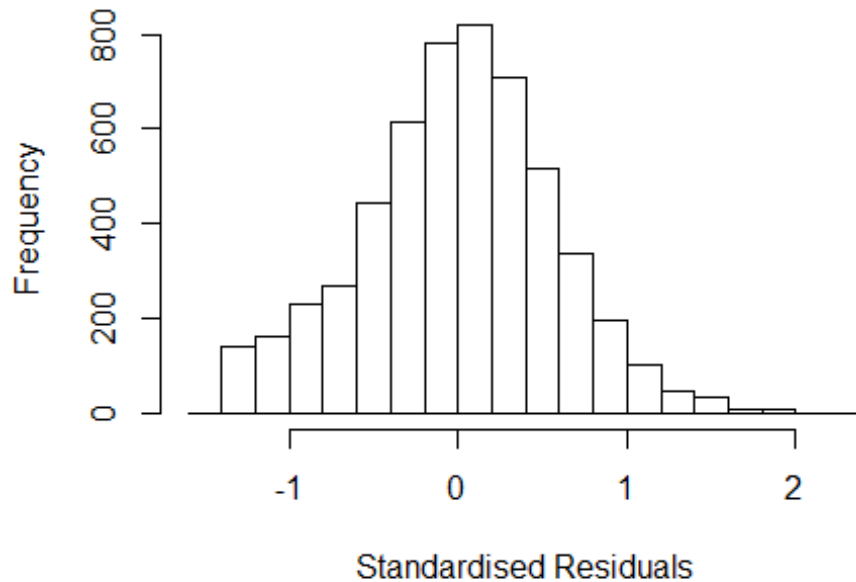
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4607 -0.3638 0.0156 0.3651 2.3415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4607 0.1127 12.96 <2e-16 ***
## FirstAuthorFemale1 0.0294 0.0212 1.39 0.1647
## Year1997 -0.2601 0.1199 -2.17 0.0301 *
## Year1998 -0.2071 0.1184 -1.75 0.0804 .
## Year1999 -0.2437 0.1164 -2.09 0.0364 *
## Year2000 -0.3319 0.1171 -2.83 0.0046 **
## Year2001 -0.2600 0.1203 -2.16 0.0308 *
## Year2002 -0.2295 0.1162 -1.97 0.0484 *
## Year2003 -0.3100 0.1172 -2.64 0.0082 **
## Year2004 -0.3239 0.1173 -2.76 0.0058 **
## Year2005 -0.3712 0.1171 -3.17 0.0015 **
## Year2006 -0.3459 0.1171 -2.95 0.0032 **
```

```

## Year2007          -0.3305      0.1165    -2.84    0.0046 **
## Year2008          -0.3351      0.1164    -2.88    0.0040 **
## Year2009          -0.2289      0.1161    -1.97    0.0488 *
## Year2010          -0.2110      0.1159    -1.82    0.0688 .
## Year2011          -0.2452      0.1162    -2.11    0.0349 *
## Year2012          -0.2173      0.1165    -1.87    0.0622 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0105, Adjusted R-squared:  0.00737
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 478 weights are ~= 1. The remaining 4932 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0207 0.8590 0.9500 0.8980 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.018 1          1.009
## Year              1.018 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4616 -0.3623 0.0149 0.3648 2.3401
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4616 0.1124 13.01 <2e-16 ***
## LastAuthorFemale1 0.0177 0.0275 0.64 0.5204
## Year1997 -0.2597 0.1196 -2.17 0.0300 *
## Year1998 -0.2061 0.1181 -1.75 0.0810 .
## Year1999 -0.2416 0.1161 -2.08 0.0375 *
## Year2000 -0.3297 0.1169 -2.82 0.0048 **
## Year2001 -0.2582 0.1201 -2.15 0.0316 *
## Year2002 -0.2285 0.1159 -1.97 0.0488 *
## Year2003 -0.3078 0.1169 -2.63 0.0085 **
## Year2004 -0.3222 0.1169 -2.75 0.0059 **
## Year2005 -0.3703 0.1168 -3.17 0.0015 **
## Year2006 -0.3432 0.1168 -2.94 0.0033 **
```

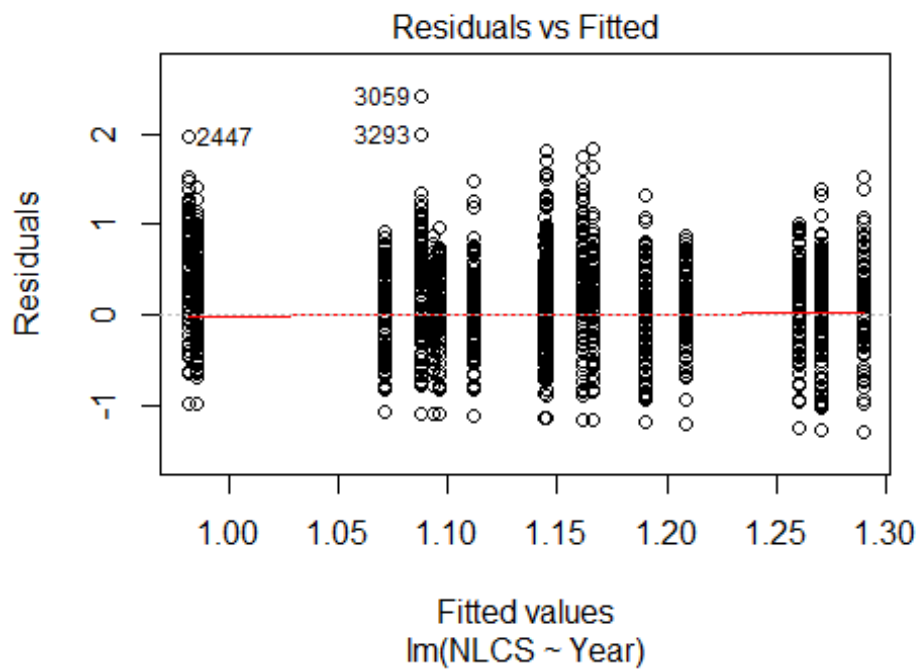


```

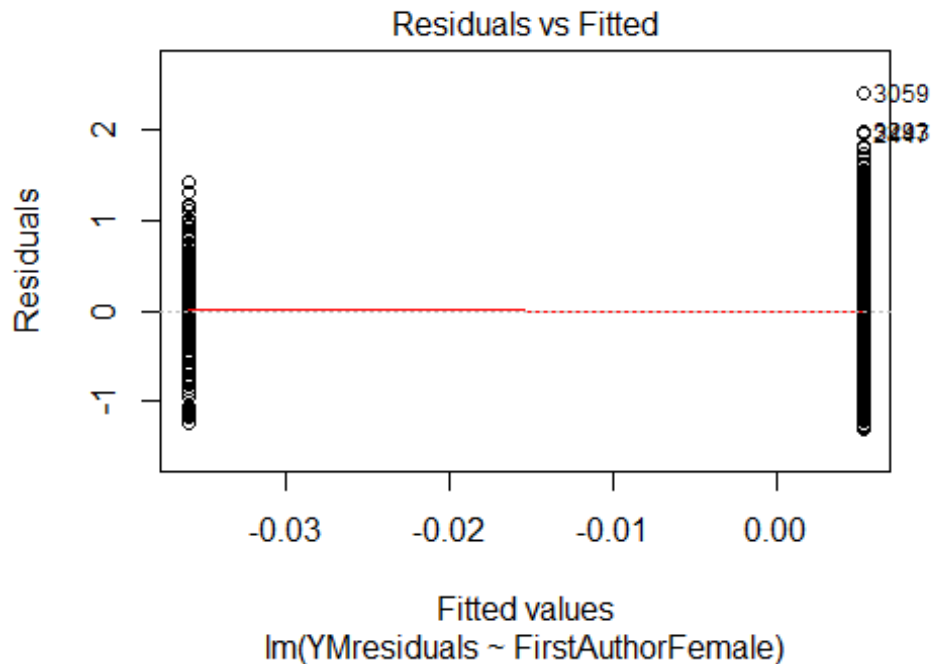
## Year2007          -0.3294      0.1161    -2.84    0.0046 **
## Year2008          -0.3339      0.1161    -2.88    0.0040 **
## Year2009          -0.2270      0.1158    -1.96    0.0501 .
## Year2010          -0.2091      0.1156    -1.81    0.0704 .
## Year2011          -0.2429      0.1159    -2.10    0.0361 *
## Year2012          -0.2155      0.1162    -1.86    0.0636 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.0102, Adjusted R-squared:  0.00712
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 476 weights are ~= 1. The remaining 4934 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0209 0.8590 0.9500 0.8980 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.85e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5410"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3101"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 348 348 342 448 437 422 428 370 437 372 407 340 324 297 247
## 2011 2012
## 297 294
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 74 95 95 125 128 154 168 117 175 111 124 120 108 108 98
## 2011 2012

```

```
## 128 120
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 64 90 92 115 119 142 151 102 157 96 103 102 90 94 89
## 2011 2012
## 110 106
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 200, df = 16, p-value <2e-16
```

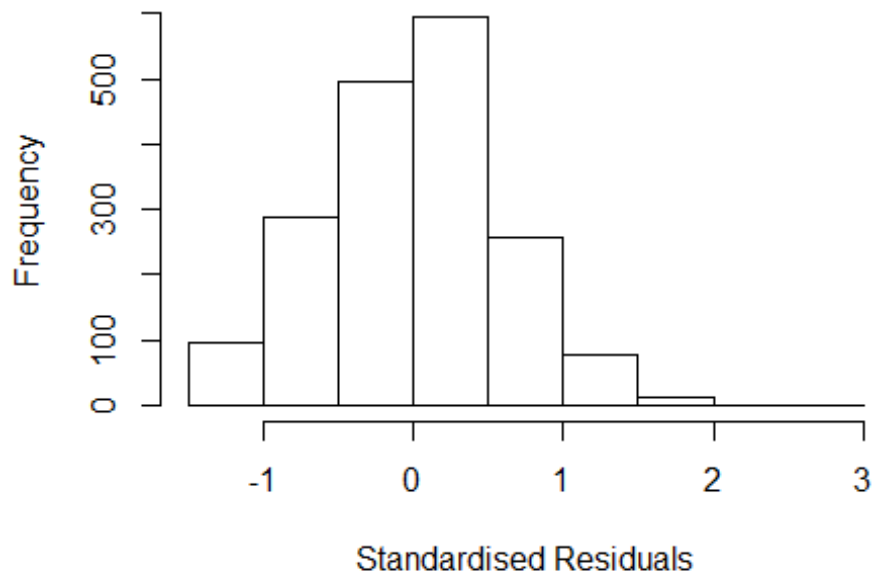


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.071, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 2.66897033513683"
## [1] "Male first author team size 2018 geometric mean: 2.15439975705311"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.6092933833503"
## [1] "Male last author team size 2018 geometric mean: 2.16583435906328"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.169 1      1.081
## LastAuthorFemale  1.167 1      1.080
## UniqueAuthors     1.228 4      1.026
## Year              1.282 16      1.008
```

## Residuals from first and last author and team size



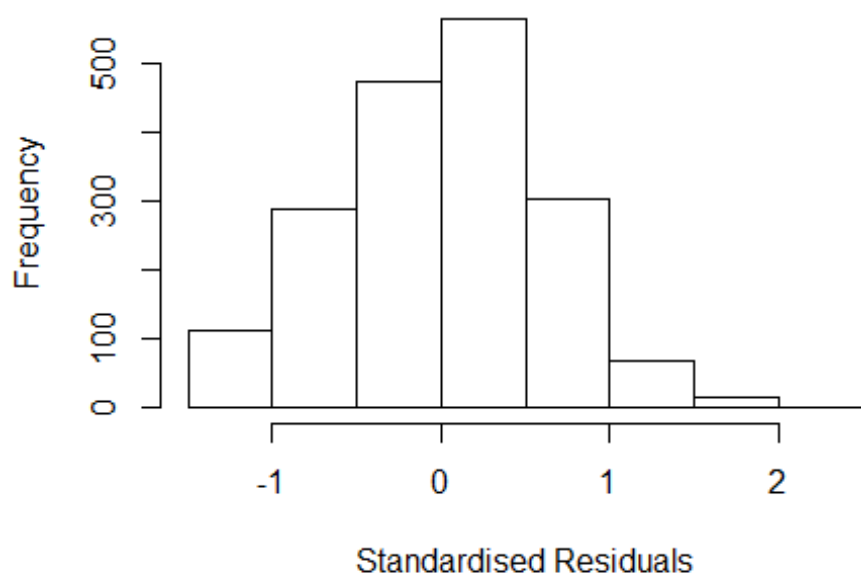
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3059 0037113847 3.5 2002    3101      2      2.558
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4539 -0.4065  0.0312  0.4016  2.5580
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1770     0.0828   14.21 < 2e-16 ***
## FirstAuthorFemale1 -0.0476     0.0469   -1.02  0.30982
## LastAuthorFemale1 -0.0265     0.0498   -0.53  0.59476
## UniqueAuthors2      0.2089     0.0371    5.63 2.1e-08 ***
## UniqueAuthors3      0.3312     0.0451    7.35 3.0e-13 ***
## UniqueAuthors4      0.2874     0.0553    5.19 2.3e-07 ***
## UniqueAuthors5      0.4091     0.0505    8.11 9.6e-16 ***
## Year1997          -0.0376     0.0996   -0.38  0.70599
## Year1998          -0.1475     0.1034   -1.43  0.15402
## Year1999          -0.1673     0.1136   -1.47  0.14089
```

```

## Year2000          -0.1322      0.1230    -1.07    0.28257
## Year2001          -0.3887      0.1145    -3.39    0.00070 ***
## Year2002          -0.2350      0.1073    -2.19    0.02858 *
## Year2003          -0.0765      0.0979    -0.78    0.43499
## Year2004          -0.3912      0.0958    -4.08    4.7e-05 ***
## Year2005          -0.1912      0.0997    -1.92    0.05538 .
## Year2006          -0.3498      0.0918    -3.81    0.00014 ***
## Year2007          -0.2443      0.0922    -2.65    0.00817 **
## Year2008          -0.2406      0.0890    -2.70    0.00695 **
## Year2009          -0.1715      0.0915    -1.87    0.06115 .
## Year2010          -0.2850      0.0936    -3.05    0.00235 **
## Year2011          -0.2242      0.0938    -2.39    0.01693 *
## Year2012          -0.3042      0.0901    -3.38    0.00075 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.586
## Multiple R-squared:  0.0755, Adjusted R-squared:  0.0642
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 145 weights are ~= 1. The remaining 1677 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.017  0.861  0.948   0.903   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.159 1          1.077
## LastAuthorFemale  1.143 1          1.069
## Year              1.062 16          1.002

```

## Residuals from first and last author



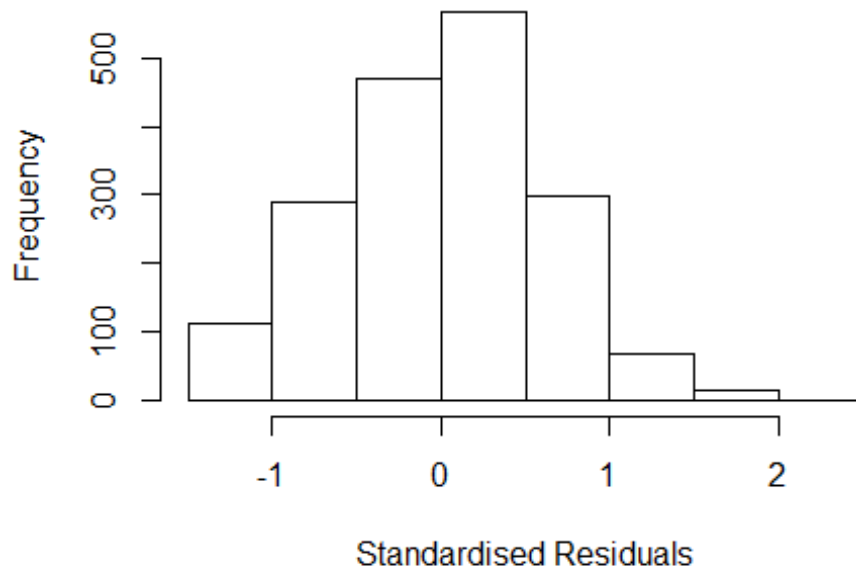
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3145 -0.4374 0.0257 0.4313 2.4021
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3145 0.0774 16.98 < 2e-16 ***
## FirstAuthorFemale1 -0.0169 0.0486 -0.35 0.72748
## LastAuthorFemale1 -0.0458 0.0529 -0.86 0.38732
## Year1997 -0.0149 0.0988 -0.15 0.88013
## Year1998 -0.1548 0.1005 -1.54 0.12368
## Year1999 -0.1883 0.1125 -1.67 0.09428 .
## Year2000 -0.1338 0.1216 -1.10 0.27104
## Year2001 -0.3540 0.1154 -3.07 0.00218 **
## Year2002 -0.2166 0.1083 -2.00 0.04554 *
## Year2003 -0.0642 0.0967 -0.66 0.50641
## Year2004 -0.3525 0.0952 -3.70 0.00022 ***
## Year2005 -0.1340 0.0971 -1.38 0.16766
```

```

## Year2006          -0.3062      0.0919   -3.33  0.00088 ***
## Year2007          -0.1773      0.0919   -1.93  0.05401 .
## Year2008          -0.1725      0.0869   -1.98  0.04737 *
## Year2009          -0.1032      0.0889   -1.16  0.24570
## Year2010          -0.2282      0.0911   -2.51  0.01233 *
## Year2011          -0.1868      0.0922   -2.03  0.04290 *
## Year2012          -0.2271      0.0890   -2.55  0.01082 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.613
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.0159
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 154 weights are ~= 1. The remaining 1668 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0896 0.8580 0.9460 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.037 1      1.018
## Year              1.037 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3110 -0.4339 0.0276 0.4286 2.4036
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3110 0.0771 17.01 < 2e-16 ***
## FirstAuthorFemale1 -0.0320 0.0463 -0.69 0.49016
## Year1997 -0.0133 0.0985 -0.14 0.89253
## Year1998 -0.1556 0.1005 -1.55 0.12171
## Year1999 -0.1891 0.1124 -1.68 0.09271 .
## Year2000 -0.1328 0.1218 -1.09 0.27561
## Year2001 -0.3534 0.1154 -3.06 0.00222 **
## Year2002 -0.2146 0.1084 -1.98 0.04802 *
## Year2003 -0.0611 0.0965 -0.63 0.52667
## Year2004 -0.3533 0.0951 -3.71 0.00021 ***
## Year2005 -0.1337 0.0971 -1.38 0.16893
## Year2006 -0.3019 0.0915 -3.30 0.00099 ***
```

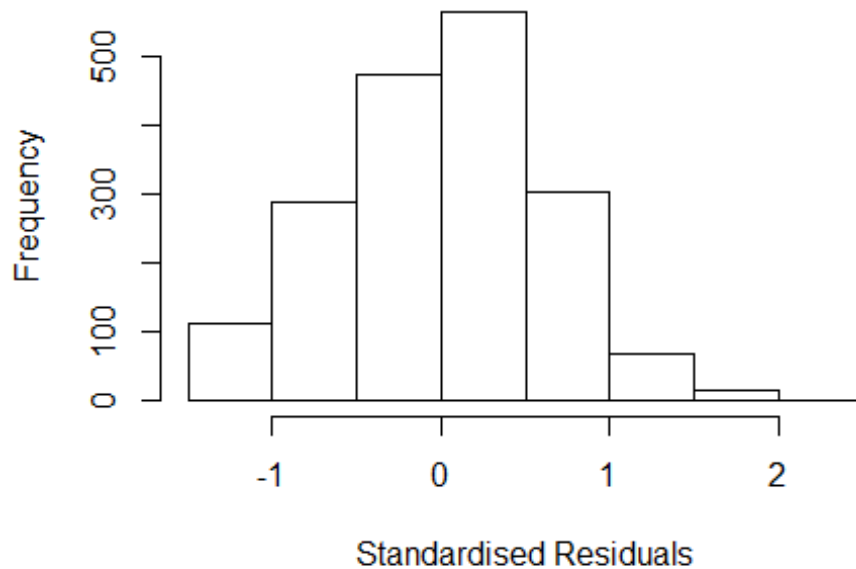


```

## Year2007          -0.1750      0.0918   -1.91  0.05675 .
## Year2008          -0.1716      0.0867   -1.98  0.04783 *
## Year2009          -0.1035      0.0886   -1.17  0.24326
## Year2010          -0.2263      0.0910   -2.49  0.01301 *
## Year2011          -0.1866      0.0922   -2.02  0.04307 *
## Year2012          -0.2273      0.0889   -2.56  0.01068 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0254, Adjusted R-squared:  0.0162
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 149 weights are ~= 1. The remaining 1673 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0824 0.8570 0.9460 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.022 1          1.011
## Year            1.022 16          1.001

```

## Residuals from last author



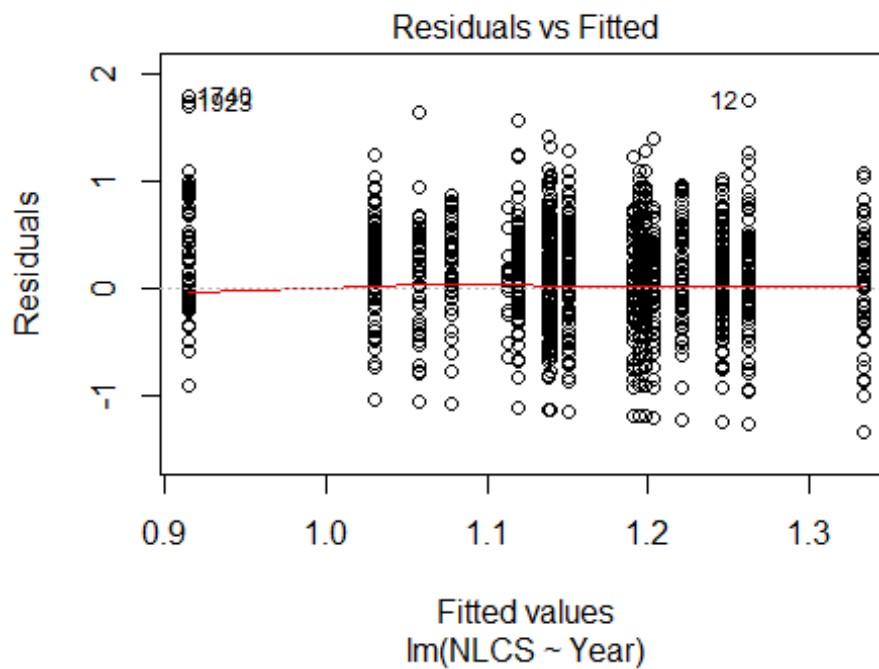
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3132 -0.4346 0.0258 0.4288 2.4035
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.3132 0.0773 16.99 < 2e-16 ***
## LastAuthorFemale1 -0.0525 0.0503 -1.04 0.29695
## Year1997 -0.0149 0.0989 -0.15 0.88006
## Year1998 -0.1537 0.1003 -1.53 0.12585
## Year1999 -0.1889 0.1126 -1.68 0.09364 .
## Year2000 -0.1342 0.1220 -1.10 0.27176
## Year2001 -0.3557 0.1157 -3.07 0.00215 **
## Year2002 -0.2168 0.1085 -2.00 0.04585 *
## Year2003 -0.0640 0.0967 -0.66 0.50795
## Year2004 -0.3530 0.0953 -3.70 0.00022 ***
## Year2005 -0.1350 0.0971 -1.39 0.16487
## Year2006 -0.3081 0.0919 -3.35 0.00081 ***
```

```

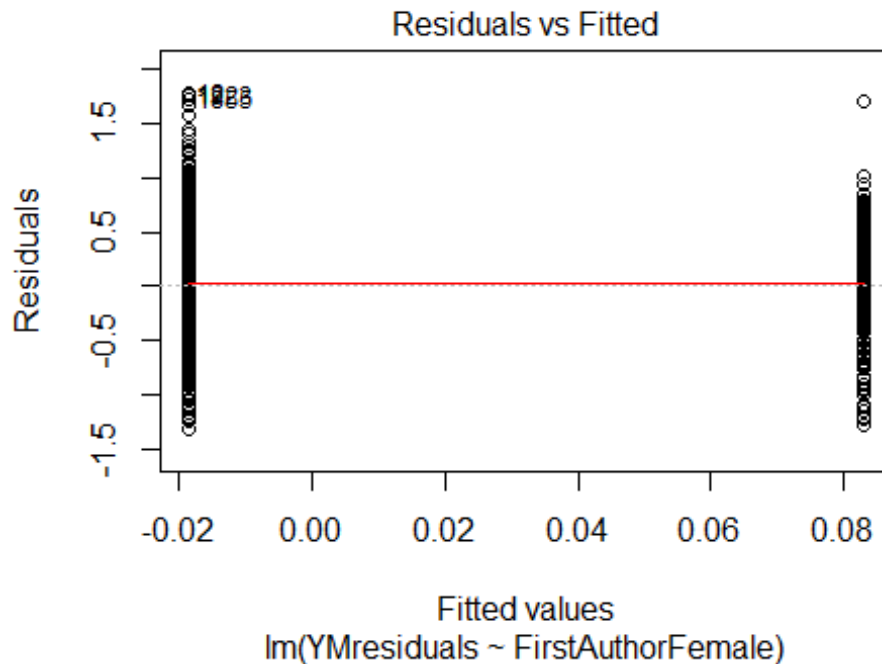
## Year2007          -0.1774      0.0920    -1.93   0.05385 .
## Year2008          -0.1721      0.0869    -1.98   0.04773 *
## Year2009          -0.1028      0.0889    -1.16   0.24748
## Year2010          -0.2277      0.0911    -2.50   0.01252 *
## Year2011          -0.1869      0.0922    -2.03   0.04279 *
## Year2012          -0.2267      0.0890    -2.55   0.01099 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0257, Adjusted R-squared:  0.0165
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 156 weights are ~= 1. The remaining 1666 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.083  0.856  0.945  0.905  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1822"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3102"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 249 215 217 203 245 184 172 248 159 196 185 226 197 215 187
## 2011 2012
## 204 157
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 78 57 64 52 61 16 60 85 53 85 88 108 83 97 103
## 2011 2012

```

```
## 116 88
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 66 48 53 47 54 15 57 73 44 71 75 98 77 86 96
## 2011 2012
## 104 79
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 2e-04
```

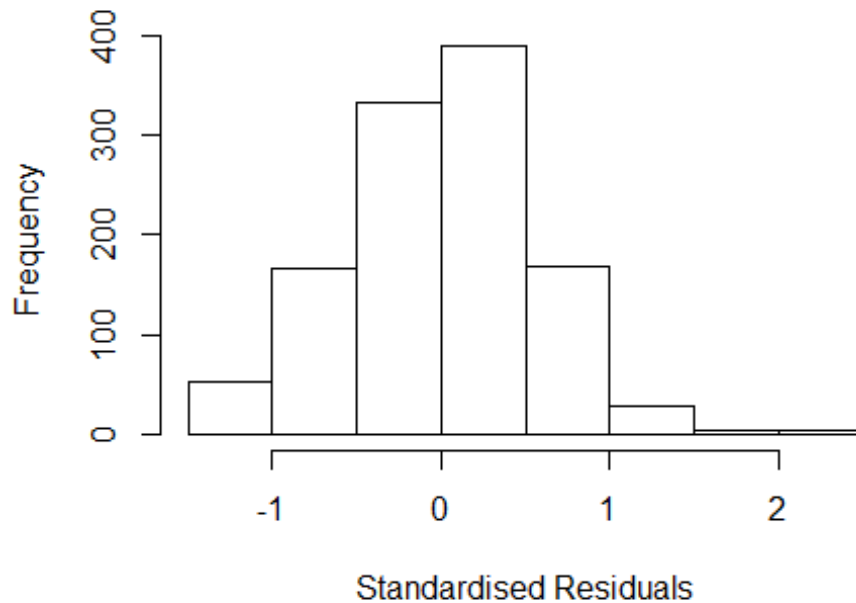


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.2, df = 1, p-value = 0.007
```



```
## [1] "Female first author team size 2018 geometric mean: 2.83585272726356"
## [1] "Male first author team size 2018 geometric mean: 2.74614917356634"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 480, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.81043779071358"
## [1] "Male last author team size 2018 geometric mean: 2.75714481387285"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 360, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.092 1          1.045
## LastAuthorFemale  1.087 1          1.042
## UniqueAuthors    1.404 4          1.043
## Year              1.472 16         1.012
```

## Residuals from first and last author and team size



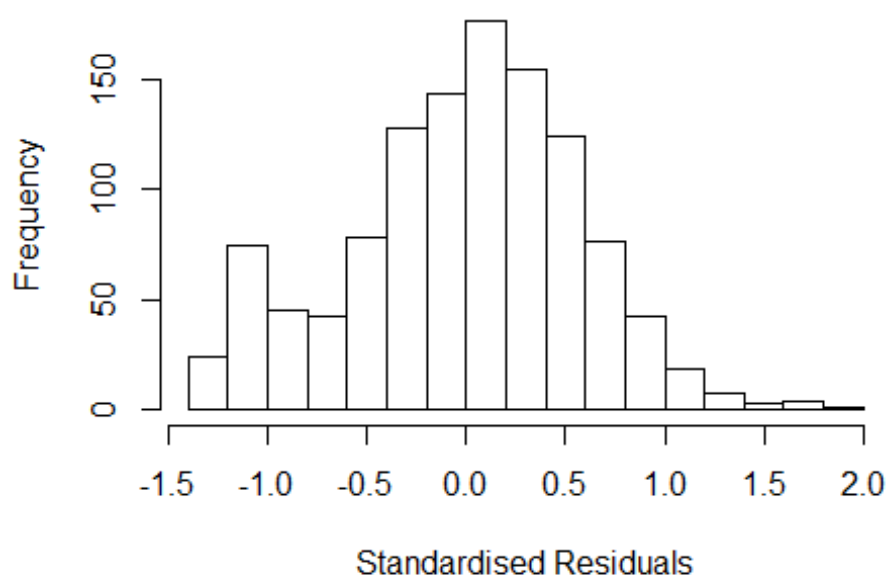
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4063 -0.3740 0.0184 0.3585 2.0828
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96149 0.08977 10.71 < 2e-16 ***
## FirstAuthorFemale1 0.05524 0.04163 1.33 0.1849
## LastAuthorFemale1 0.01182 0.05625 0.21 0.8335
## UniqueAuthors2 0.43055 0.05278 8.16 9.1e-16 ***
## UniqueAuthors3 0.42519 0.05471 7.77 1.7e-14 ***
## UniqueAuthors4 0.43132 0.06618 6.52 1.1e-10 ***
## UniqueAuthors5 0.48930 0.06270 7.80 1.4e-14 ***
## Year1997 0.02525 0.11308 0.22 0.8234
## Year1998 -0.06568 0.10740 -0.61 0.5410
## Year1999 -0.27468 0.11677 -2.35 0.0188 *
```

```

## Year2000          0.00244    0.12524    0.02    0.9844
## Year2001          -0.21477    0.13748   -1.56    0.1185
## Year2002          -0.11821    0.11079   -1.07    0.2862
## Year2003          -0.38234    0.12699   -3.01    0.0027 **
## Year2004          -0.22541    0.11816   -1.91    0.0567 .
## Year2005          -0.28730    0.11742   -2.45    0.0146 *
## Year2006          -0.17638    0.11014   -1.60    0.1096
## Year2007          -0.14328    0.10263   -1.40    0.1630
## Year2008          -0.09885    0.11139   -0.89    0.3750
## Year2009          -0.17425    0.10784   -1.62    0.1064
## Year2010          -0.19343    0.10154   -1.91    0.0570 .
## Year2011          -0.11244    0.09639   -1.17    0.2436
## Year2012          -0.16300    0.11249   -1.45    0.1476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.131, Adjusted R-squared:  0.114
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1037 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.857  0.950  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.75e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.066 1          1.033
## LastAuthorFemale 1.073 1          1.036
## Year          1.104 16          1.003

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3412 -0.3620  0.0461  0.3832  1.8120
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.21202    0.08811   13.76  <2e-16 ***
## FirstAuthorFemale1 0.11858    0.04242    2.80  0.0053 **
## LastAuthorFemale1 0.02001    0.05735    0.35  0.7272
## Year1997         0.12922    0.11658    1.11  0.2679
## Year1998        -0.03353    0.10684   -0.31  0.7537
## Year1999        -0.18254    0.12912   -1.41  0.1577
## Year2000         0.03316    0.12188    0.27  0.7856
## Year2001        -0.14662    0.13631   -1.08  0.2823
## Year2002        -0.04861    0.11029   -0.44  0.6595
## Year2003        -0.33627    0.13402   -2.51  0.0122 *
## Year2004        -0.15984    0.12757   -1.25  0.2105
## Year2005        -0.20738    0.12348   -1.68  0.0933 .
```

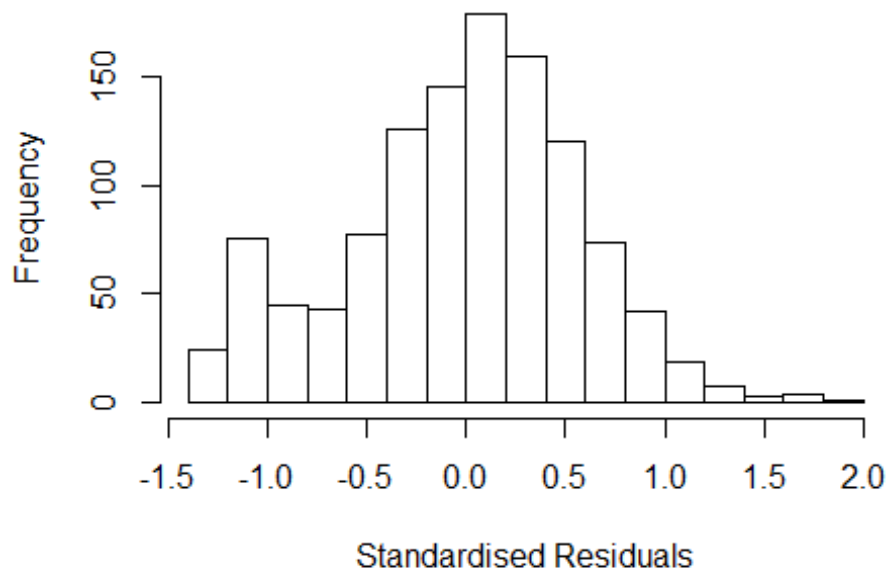


```

## Year2006      -0.12894    0.11876   -1.09    0.2778
## Year2007      -0.05614    0.10617   -0.53    0.5970
## Year2008      -0.01376    0.11477   -0.12    0.9046
## Year2009      -0.05155    0.11072   -0.47    0.6416
## Year2010      -0.10077    0.10444   -0.96    0.3348
## Year2011       0.00268    0.09731    0.03    0.9781
## Year2012      -0.07108    0.10999   -0.65    0.5182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0342, Adjusted R-squared:  0.0187
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 1047 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.258  0.853   0.951   0.897   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.047 1      1.023
## Year              1.047 16      1.001

```

## Residuals from first author



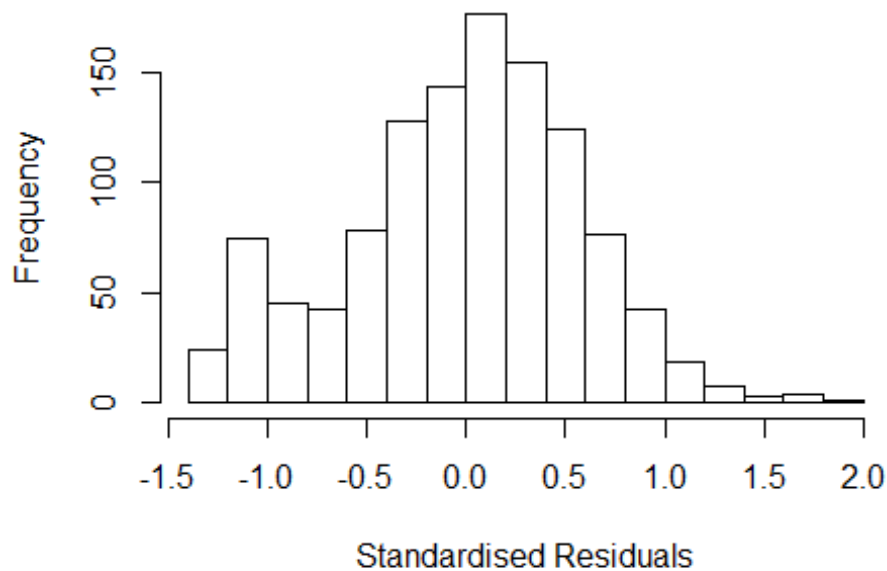
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3426 -0.3640 0.0449 0.3847 1.8101
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2139 0.0876 13.85 <2e-16 ***
## FirstAuthorFemale1 0.1216 0.0421 2.89 0.0039 **
## Year1997 0.1287 0.1167 1.10 0.2703
## Year1998 -0.0341 0.1068 -0.32 0.7496
## Year1999 -0.1811 0.1298 -1.40 0.1630
## Year2000 0.0332 0.1217 0.27 0.7849
## Year2001 -0.1478 0.1361 -1.09 0.2777
## Year2002 -0.0497 0.1101 -0.45 0.6516
## Year2003 -0.3366 0.1339 -2.51 0.0121 *
## Year2004 -0.1599 0.1276 -1.25 0.2103
## Year2005 -0.2081 0.1234 -1.69 0.0920 .
## Year2006 -0.1285 0.1188 -1.08 0.2798
```

```

## Year2007          -0.0563      0.1062    -0.53    0.5959
## Year2008          -0.0141      0.1147    -0.12    0.9022
## Year2009          -0.0532      0.1105    -0.48    0.6302
## Year2010          -0.1008      0.1045    -0.96    0.3348
## Year2011           0.0035      0.0974     0.04    0.9714
## Year2012          -0.0703      0.1100    -0.64    0.5233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0341, Adjusted R-squared:  0.0195
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 1047 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.258  0.852  0.950  0.896  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.056 1          1.028
## Year            1.056 16          1.002

```

## Residuals from last author



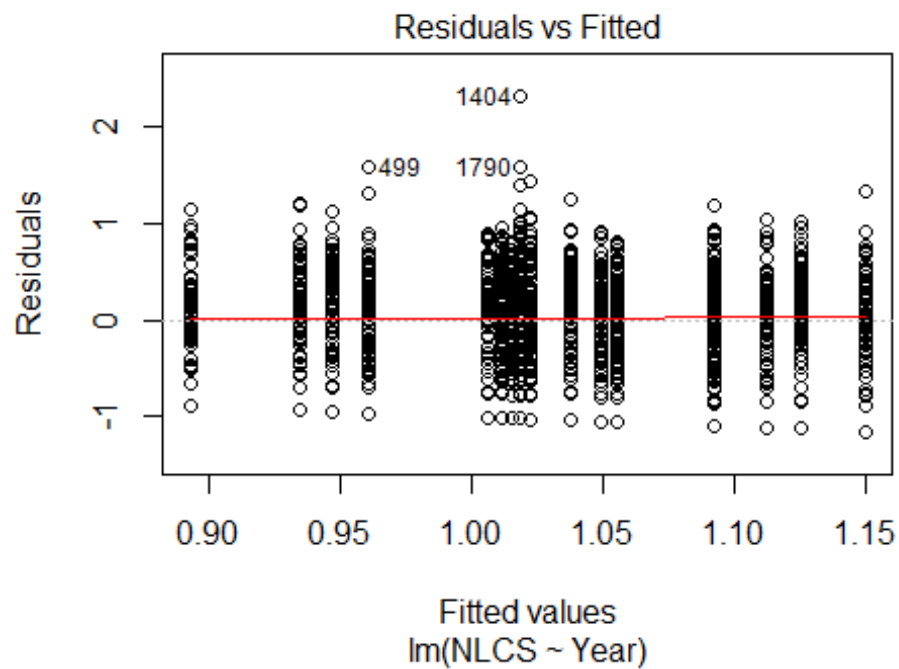
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3552 -0.3659  0.0422  0.3921  1.8202
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.22183    0.08825   13.85  <2e-16 ***
## LastAuthorFemale1 0.04727    0.05596    0.84   0.398
## Year1997        0.13335    0.11727    1.14   0.256
## Year1998       -0.03281    0.10746   -0.31   0.760
## Year1999       -0.18274    0.12917   -1.41   0.157
## Year2000        0.03795    0.12311    0.31   0.758
## Year2001       -0.11375    0.13368   -0.85   0.395
## Year2002       -0.03421    0.10972   -0.31   0.755
## Year2003       -0.33306    0.13583   -2.45   0.014 *
## Year2004       -0.14562    0.13006   -1.12   0.263
## Year2005       -0.19074    0.12448   -1.53   0.126
## Year2006       -0.10438    0.11825   -0.88   0.378
```

```

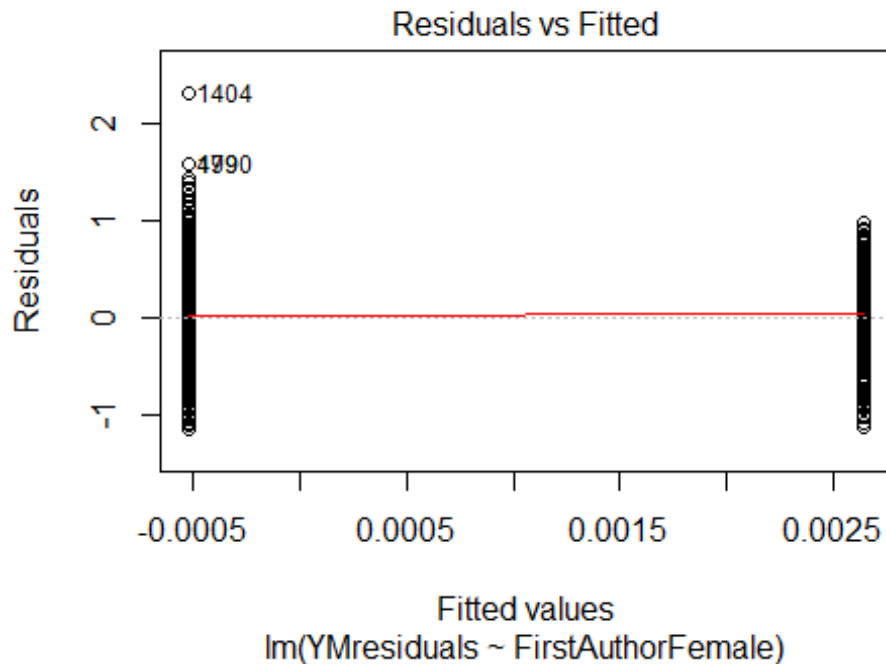
## Year2007          -0.04654      0.10686   -0.44      0.663
## Year2008          -0.00542      0.11561   -0.05      0.963
## Year2009          -0.03330      0.11058   -0.30      0.763
## Year2010          -0.09376      0.10471   -0.90      0.371
## Year2011           0.01042      0.09782    0.11      0.915
## Year2012          -0.05935      0.11019   -0.54      0.590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.551
## Multiple R-squared:  0.0285, Adjusted R-squared:  0.0138
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 94 weights are ~= 1. The remaining 1049 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.252  0.848  0.949  0.896  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.75e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1143"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3103"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 445 404 419 382 349 329 354 301 301 313 328 317 309 410 386
## 2011 2012
## 457 395
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 115 123 132 101 81 79 97 96 88 114 92 107 96 143 135
## 2011 2012

```

```
## 174 160
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 99 109 117 91 77 73 81 84 84 102 86 98 83 133 121
## 2011 2012
## 156 147
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 35, df = 16, p-value = 0.004
```

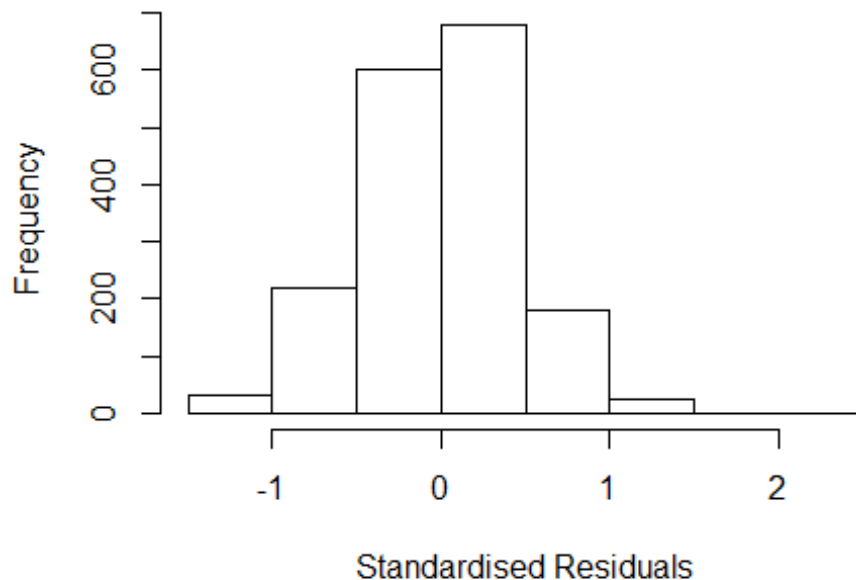


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.68114812664161"
## [1] "Male first author team size 2018 geometric mean: 2.13486989204082"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3500, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.51006275963741"
## [1] "Male last author team size 2018 geometric mean: 2.18338708473846"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.088 1      1.043
## LastAuthorFemale  1.054 1      1.027
## UniqueAuthors     1.284 4      1.032
## Year              1.348 16     1.009
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3349 -0.2814 0.0126 0.3029 2.1478
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.77759 0.04920 15.80 < 2e-16 ***
## FirstAuthorFemale1 -0.02472 0.02873 -0.86 0.390
## LastAuthorFemale1 -0.05356 0.03449 -1.55 0.121
## UniqueAuthors2 0.31705 0.03003 10.56 < 2e-16 ***
## UniqueAuthors3 0.39541 0.03253 12.16 < 2e-16 ***
## UniqueAuthors4 0.35955 0.04500 7.99 2.5e-15 ***
## UniqueAuthors5 0.40306 0.05095 7.91 4.5e-15 ***
## Year1997 0.06603 0.07054 0.94 0.349
## Year1998 0.01124 0.07153 0.16 0.875
## Year1999 0.13020 0.06199 2.10 0.036 *
```

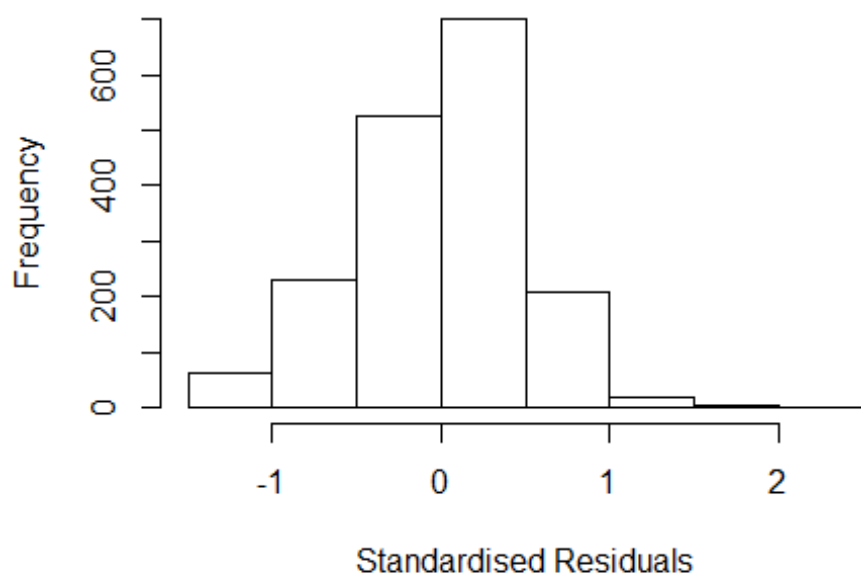


```

## Year2000          0.16192      0.07468      2.17      0.030 *
## Year2001          0.05821      0.07126      0.82      0.414
## Year2002          0.01659      0.06717      0.25      0.805
## Year2003         -0.05953      0.07402     -0.80      0.421
## Year2004          0.09698      0.06810      1.42      0.155
## Year2005          0.00894      0.07309      0.12      0.903
## Year2006          0.00482      0.06944      0.07      0.945
## Year2007         -0.05740      0.07289     -0.79      0.431
## Year2008         -0.15875      0.06830     -2.32      0.020 *
## Year2009          0.10732      0.06154      1.74      0.081 .
## Year2010          0.06150      0.05951      1.03      0.302
## Year2011          0.01286      0.05833      0.22      0.826
## Year2012          0.12586      0.06093      2.07      0.039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.145, Adjusted R-squared:  0.134
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 223 is an outlier with |weight| = 0 ( < 5.7e-05);
## 146 weights are ~1. The remaining 1594 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.226  0.863   0.951   0.901   0.986   0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           5.74e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.044 1           1.022
## LastAuthorFemale  1.046 1           1.023
## Year              1.069 16           1.002

```

## Residuals from first and last author



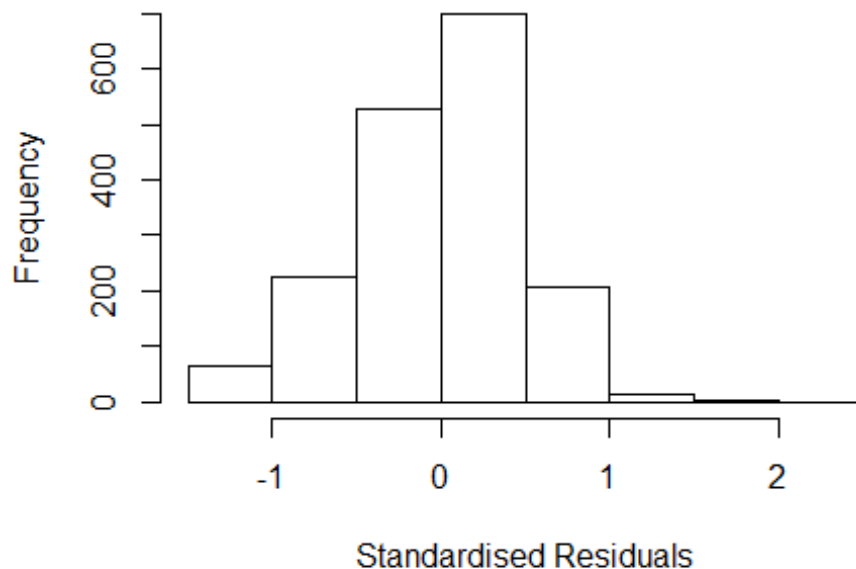
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1974 -0.3321 0.0369 0.3273 2.3598
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9321 0.0516 18.08 < 2e-16 ***
## FirstAuthorFemale1 0.0333 0.0303 1.10 0.27261
## LastAuthorFemale1 -0.0564 0.0379 -1.49 0.13674
## Year1997 0.0757 0.0741 1.02 0.30738
## Year1998 0.0401 0.0765 0.52 0.60004
## Year1999 0.1989 0.0672 2.96 0.00314 **
## Year2000 0.2251 0.0766 2.94 0.00335 **
## Year2001 0.1318 0.0801 1.65 0.09982 .
## Year2002 0.0932 0.0728 1.28 0.20026
## Year2003 -0.0240 0.0784 -0.31 0.76006
## Year2004 0.1650 0.0747 2.21 0.02739 *
## Year2005 0.1053 0.0782 1.35 0.17834
```

```

## Year2006          0.0840      0.0729      1.15  0.24974
## Year2007          0.0215      0.0791      0.27  0.78588
## Year2008         -0.0664      0.0734     -0.90  0.36562
## Year2009          0.2009      0.0644      3.12  0.00183 **
## Year2010          0.1650      0.0633      2.61  0.00917 **
## Year2011          0.1016      0.0633      1.60  0.10888
## Year2012          0.2321      0.0627      3.70  0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0313, Adjusted R-squared:  0.0212
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 223 is an outlier with |weight| = 0 ( < 5.7e-05);
## 146 weights are ~= 1. The remaining 1594 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.197  0.866   0.948   0.900   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.032 1          1.016
## Year              1.032 16          1.001

```

## Residuals from first author



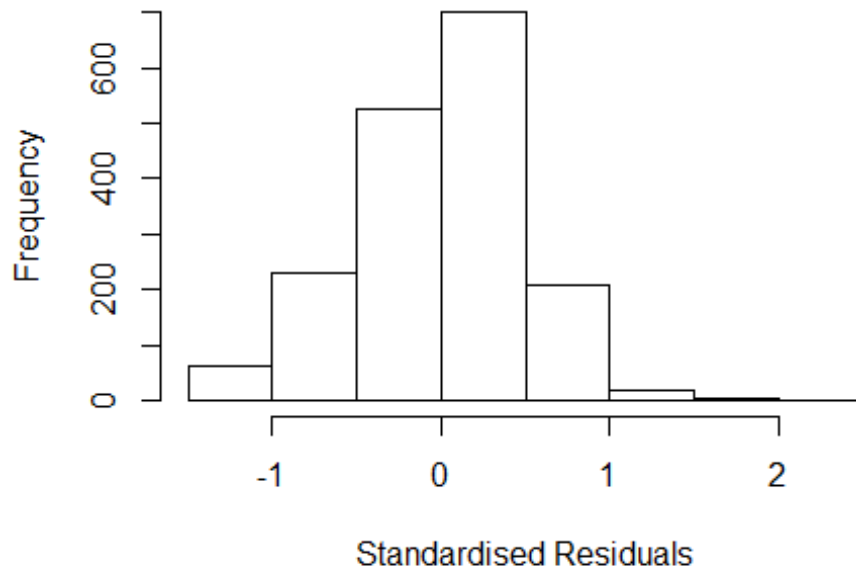
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1791 -0.3316 0.0356 0.3247 2.3610
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9286 0.0515 18.02 < 2e-16 ***
## FirstAuthorFemale1 0.0208 0.0311 0.67 0.50409
## Year1997 0.0720 0.0743 0.97 0.33271
## Year1998 0.0424 0.0765 0.55 0.57926
## Year1999 0.1942 0.0673 2.88 0.00397 **
## Year2000 0.2293 0.0765 3.00 0.00278 **
## Year2001 0.1321 0.0798 1.65 0.09826 .
## Year2002 0.0936 0.0730 1.28 0.19995
## Year2003 -0.0250 0.0784 -0.32 0.74945
## Year2004 0.1675 0.0750 2.23 0.02565 *
## Year2005 0.1049 0.0784 1.34 0.18086
## Year2006 0.0857 0.0732 1.17 0.24199
```

```

## Year2007          0.0197      0.0796      0.25  0.80480
## Year2008         -0.0711      0.0733     -0.97  0.33211
## Year2009          0.2009      0.0644      3.12  0.00186 **
## Year2010          0.1637      0.0634      2.58  0.00995 **
## Year2011          0.1001      0.0632      1.58  0.11357
## Year2012          0.2298      0.0626      3.67  0.00025 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0303, Adjusted R-squared:  0.0207
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 223 is an outlier with |weight| = 0 ( < 5.7e-05);
## 149 weights are ~= 1. The remaining 1591 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.196  0.863  0.948  0.900  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.035 1          1.018
## Year            1.035 16          1.001

```

## Residuals from last author



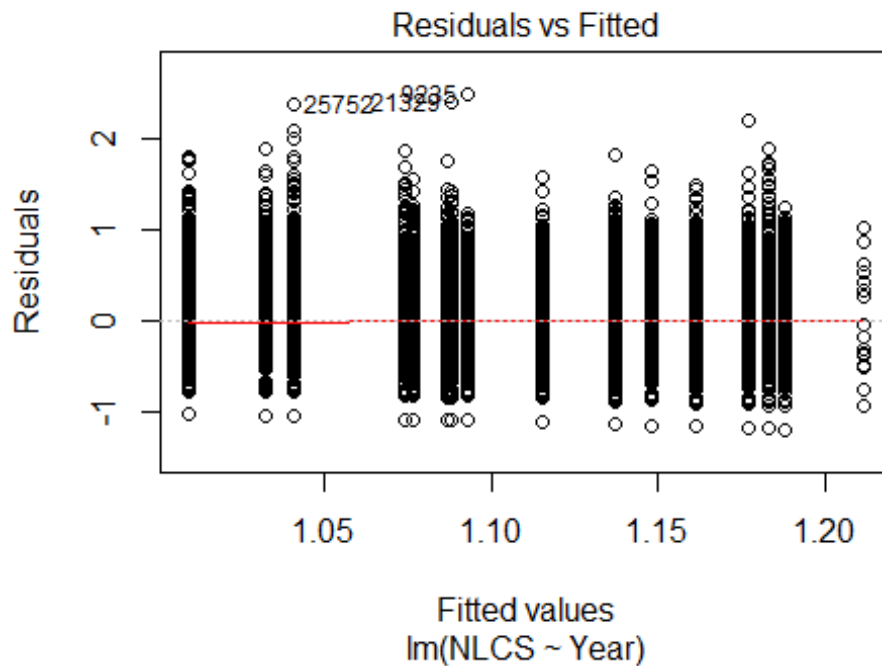
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1665 -0.3304 0.0334 0.3253 2.3562
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9381 0.0514 18.25 < 2e-16 ***
## LastAuthorFemale1 -0.0459 0.0377 -1.22 0.22354
## Year1997 0.0746 0.0745 1.00 0.31679
## Year1998 0.0377 0.0769 0.49 0.62380
## Year1999 0.1964 0.0676 2.91 0.00371 **
## Year2000 0.2255 0.0769 2.93 0.00342 **
## Year2001 0.1286 0.0803 1.60 0.10937
## Year2002 0.0920 0.0731 1.26 0.20841
## Year2003 -0.0259 0.0786 -0.33 0.74157
## Year2004 0.1621 0.0749 2.17 0.03048 *
## Year2005 0.1043 0.0786 1.33 0.18461
## Year2006 0.0815 0.0733 1.11 0.26677
```

```

## Year2007          0.0213      0.0794      0.27  0.78864
## Year2008         -0.0675      0.0736     -0.92  0.35939
## Year2009          0.1989      0.0645      3.08  0.00208 **
## Year2010          0.1648      0.0637      2.59  0.00972 **
## Year2011          0.0998      0.0635      1.57  0.11625
## Year2012          0.2284      0.0627      3.64  0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.466
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.0212
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 223 is an outlier with |weight| = 0 ( < 5.7e-05);
## 149 weights are ~= 1. The remaining 1591 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.199  0.865   0.948   0.900   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1741"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3104"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2193 2118 1944 158 2099 2019 1761 1559 1524 1554 1641 1608 1582 1777 1680
## 2011 2012
## 1710 1590
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 367 395 361 18 367 423 532 497 511 559 583 600 607 630 650

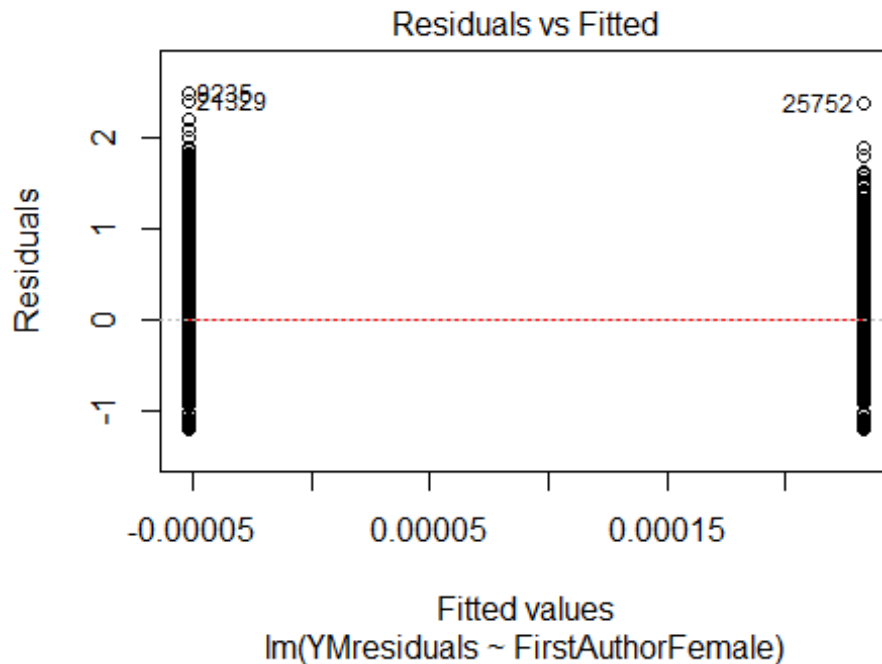
```

```
## 2011 2012
## 721 660
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 326 344 305 16 323 368 431 443 434 475 500 504 519 525 562
## 2011 2012
## 630 548
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 98, df = 16, p-value = 1e-13
```



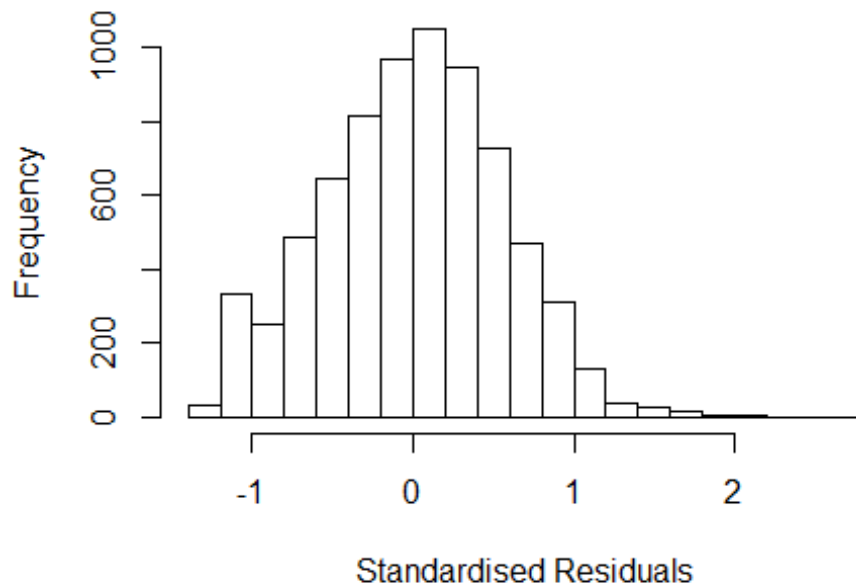
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.9, df = 1, p-value = 0.2
```





```
## [1] "Female first author team size 2018 geometric mean: 3.69518869192319"
## [1] "Male first author team size 2018 geometric mean: 2.98243997519324"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34000, p-value = 5e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.55848561171493"
## [1] "Male last author team size 2018 geometric mean: 3.06012051432474"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 20000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.046 1 1.023
## LastAuthorFemale 1.023 1 1.011
## UniqueAuthors 1.093 4 1.011
## Year 1.086 16 1.003
```

## Residuals from first and last author and team size



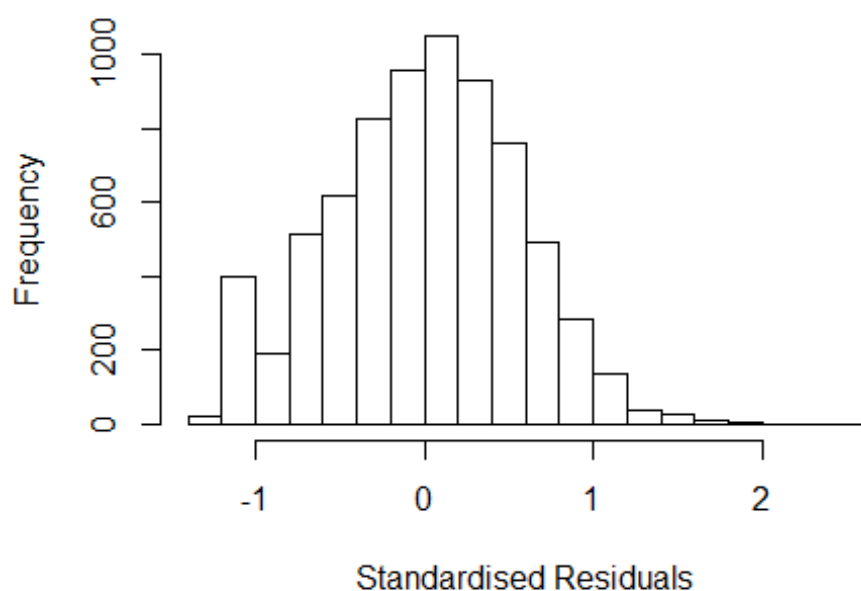
```
## [1] "List of 2 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 9235    0034187380 3.580 2000    2208      2    2.611
## 25752 34249889935 3.412 2007    1600      3    2.521
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2765 -0.3795  0.0169  0.3817  2.6112
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9068    0.0396   22.89 < 2e-16 ***
## FirstAuthorFemale1 -0.0110    0.0173   -0.64  0.52400
## LastAuthorFemale1  0.0120    0.0242    0.49  0.62090
## UniqueAuthors2    0.1314    0.0223    5.90  3.7e-09 ***
## UniqueAuthors3    0.1274    0.0231    5.52  3.5e-08 ***
## UniqueAuthors4    0.1533    0.0259    5.92  3.4e-09 ***
## UniqueAuthors5    0.2002    0.0245    8.16  4.0e-16 ***
## Year1997          0.1091    0.0481    2.27  0.02348 *
## Year1998          0.1427    0.0497    2.87  0.00411 **
```

```

## Year1999          0.1105      0.1614      0.68  0.49389
## Year2000          0.0620      0.0495      1.25  0.21036
## Year2001          0.0134      0.0536      0.25  0.80248
## Year2002          0.0667      0.0478      1.40  0.16239
## Year2003         -0.0392      0.0488     -0.80  0.42206
## Year2004          0.0585      0.0460      1.27  0.20299
## Year2005          0.0259      0.0451      0.57  0.56532
## Year2006          0.0344      0.0448      0.77  0.44167
## Year2007         -0.0163      0.0464     -0.35  0.72526
## Year2008          0.1264      0.0444      2.85  0.00443 **
## Year2009          0.1354      0.0446      3.03  0.00243 **
## Year2010          0.1694      0.0431      3.93  8.6e-05 ***
## Year2011          0.1513      0.0429      3.53  0.00042 ***
## Year2012          0.0979      0.0432      2.27  0.02350 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.561
## Multiple R-squared:  0.0256, Adjusted R-squared:  0.0226
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 591 weights are ~= 1. The remaining 6662 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0002 0.8700 0.9500 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.012
## LastAuthorFemale  1.018 1      1.009
## Year              1.022 16      1.001

```

## Residuals from first and last author



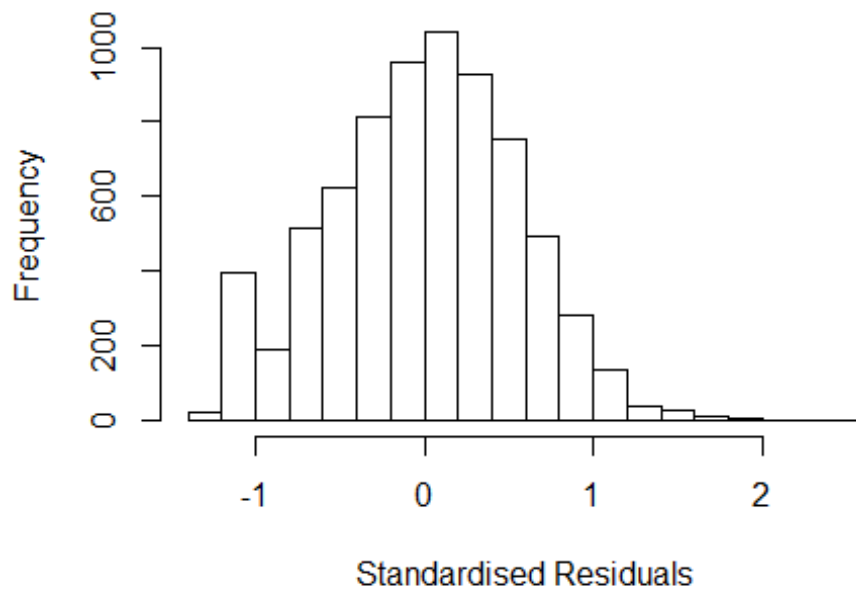
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 9235 0034187380 3.58 2000      2208      2      2.503
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.226 -0.376  0.019  0.387  2.503
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.00256    0.03777   26.55 < 2e-16 ***
## FirstAuthorFemale1 0.00935    0.01714    0.55  0.58558
## LastAuthorFemale1 0.01544    0.02423    0.64  0.52404
## Year1997          0.10869    0.04897    2.22  0.02649 *
## Year1998          0.15293    0.05033    3.04  0.00239 **
## Year1999          0.14722    0.16256    0.91  0.36518
## Year2000          0.07428    0.05000    1.49  0.13748
## Year2001          0.04254    0.05400    0.79  0.43085
## Year2002          0.08345    0.04795    1.74  0.08179 .
## Year2003         -0.02254    0.04939   -0.46  0.64812
## Year2004          0.08207    0.04628    1.77  0.07620 .
## Year2005          0.05050    0.04562    1.11  0.26826
```

```

## Year2006          0.06073      0.04520      1.34  0.17917
## Year2007          0.01110      0.04665      0.24  0.81196
## Year2008          0.15378      0.04490      3.42  0.00062 ***
## Year2009          0.16205      0.04508      3.59  0.00033 ***
## Year2010          0.19886      0.04351      4.57  4.9e-06 ***
## Year2011          0.18512      0.04318      4.29  1.8e-05 ***
## Year2012          0.13438      0.04366      3.08  0.00209 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.014, Adjusted R-squared:  0.0115
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 606 weights are ~= 1. The remaining 6647 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0114 0.8700 0.9500 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.013 1      1.007
## Year              1.013 16      1.000

```

## Residuals from first author



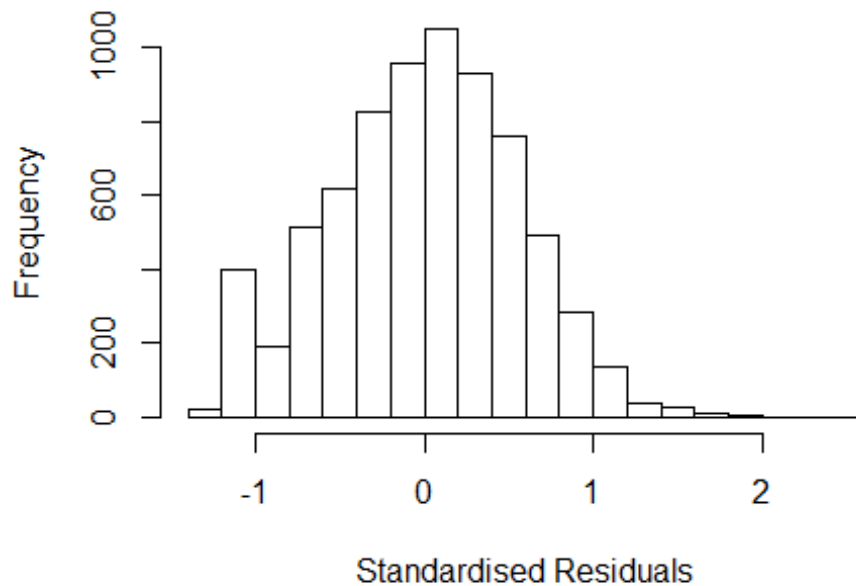
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 9235 0034187380 3.58 2000      2208      2      2.503
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2133 -0.3760  0.0186  0.3874  2.5019
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0034     0.0377   26.59 < 2e-16 ***
## FirstAuthorFemale1 0.0108     0.0171    0.63  0.52643
## Year1997          0.1087     0.0490    2.22  0.02651 *
## Year1998          0.1530     0.0503    3.04  0.00237 **
## Year1999          0.1485     0.1626    0.91  0.36119
## Year2000          0.0747     0.0500    1.49  0.13539
## Year2001          0.0428     0.0540    0.79  0.42780
## Year2002          0.0842     0.0479    1.76  0.07900 .
## Year2003         -0.0222     0.0494   -0.45  0.65378
## Year2004          0.0822     0.0463    1.78  0.07563 .
## Year2005          0.0505     0.0456    1.11  0.26811
## Year2006          0.0612     0.0452    1.35  0.17584
```

```

## Year2007          0.0117      0.0466      0.25  0.80158
## Year2008          0.1544      0.0449      3.44  0.00059 ***
## Year2009          0.1623      0.0451      3.60  0.00032 ***
## Year2010          0.1991      0.0435      4.58  4.8e-06 ***
## Year2011          0.1854      0.0432      4.29  1.8e-05 ***
## Year2012          0.1346      0.0437      3.08  0.00206 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.0139, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 604 weights are ~= 1. The remaining 6649 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0115 0.8710 0.9500 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.009 1      1.004
## Year              1.009 16      1.000

```

## Residuals from last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 9235 0034187380 3.58 2000      2208      2      2.503
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2203 -0.3767  0.0183  0.3860  2.5016
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0036     0.0377  26.62 < 2e-16 ***
## LastAuthorFemale1  0.0170     0.0241   0.71  0.47988
## Year1997          0.1087     0.0490   2.22  0.02651 *
## Year1998          0.1531     0.0503   3.04  0.00236 **
## Year1999          0.1473     0.1628   0.90  0.36567
## Year2000          0.0748     0.0500   1.50  0.13477
## Year2001          0.0427     0.0540   0.79  0.42884
## Year2002          0.0833     0.0479   1.74  0.08218 .
## Year2003         -0.0220     0.0494  -0.45  0.65529
## Year2004          0.0824     0.0463   1.78  0.07479 .
## Year2005          0.0513     0.0456   1.13  0.26055
## Year2006          0.0616     0.0452   1.37  0.17229
```

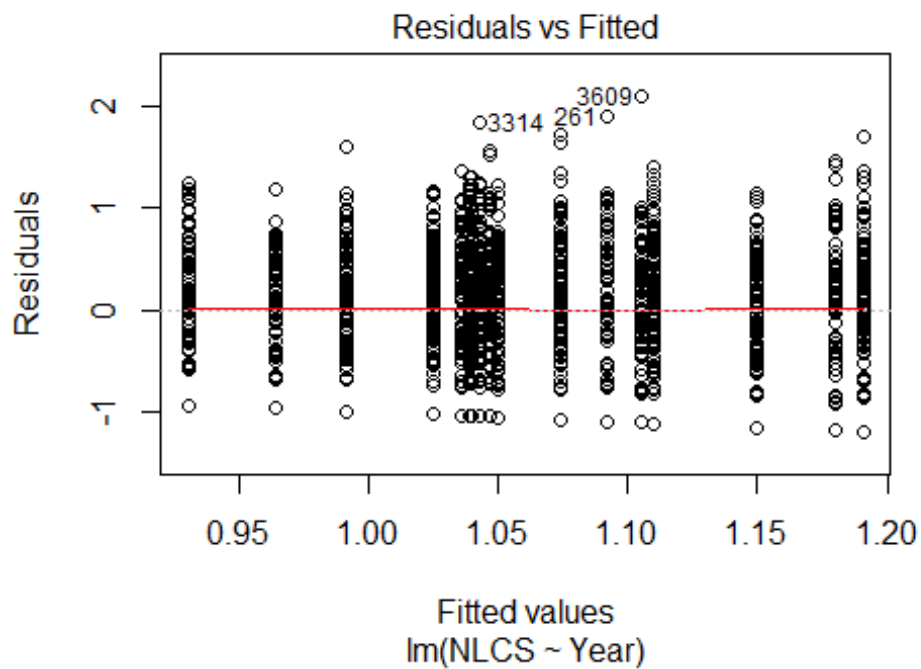


```

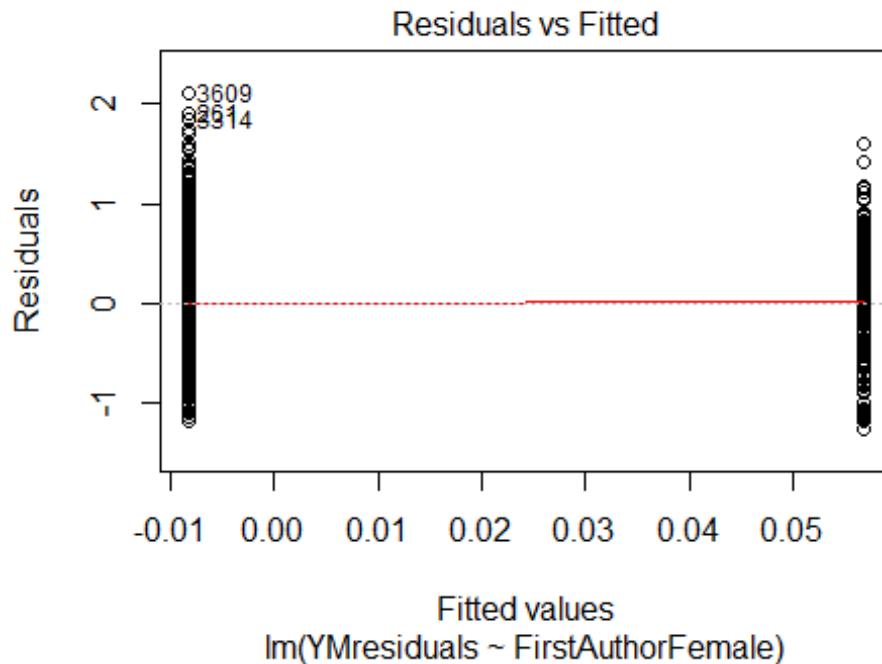
## Year2007          0.0122      0.0466      0.26  0.79378
## Year2008          0.1545      0.0449      3.44  0.00058 ***
## Year2009          0.1625      0.0451      3.61  0.00031 ***
## Year2010          0.1997      0.0435      4.59  4.4e-06 ***
## Year2011          0.1857      0.0432      4.30  1.7e-05 ***
## Year2012          0.1351      0.0436      3.10  0.00196 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.565
## Multiple R-squared:  0.014, Adjusted R-squared:  0.0116
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 609 weights are ~= 1. The remaining 6644 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0116 0.8700 0.9500 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.38e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7253"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3105"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 399 485 386 427 443 348 349 245 272 256 302 330 304 360 264
## 2011 2012
## 291 290
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 84 113 97 89 115 78 111 97 100 101 115 134 116 138 107
## 2011 2012

```

```
## 122 108
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 96 92 80 103 67 96 85 91 96 95 108 105 118 94
## 2011 2012
## 101 99
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 28, df = 16, p-value = 0.03
```

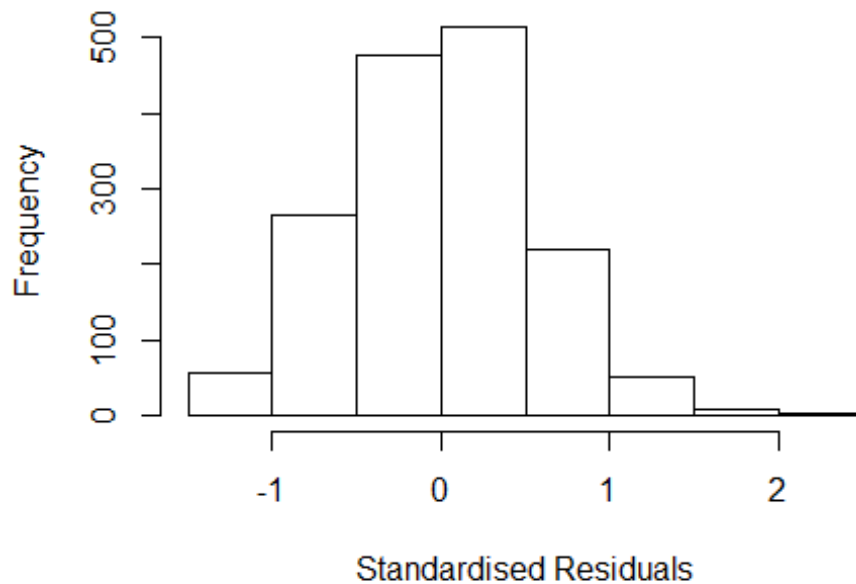


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.15, df = 1, p-value = 0.7
```



```
## [1] "Female first author team size 2018 geometric mean: 3.9629149634319"
## [1] "Male first author team size 2018 geometric mean: 3.20218029378125"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3100, p-value = 0.06
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.59515761454964"
## [1] "Male last author team size 2018 geometric mean: 3.29296068706675"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.162 1      1.078
## LastAuthorFemale  1.148 1      1.071
## UniqueAuthors    1.210 4      1.024
## Year              1.228 16     1.006
```

## Residuals from first and last author and team size



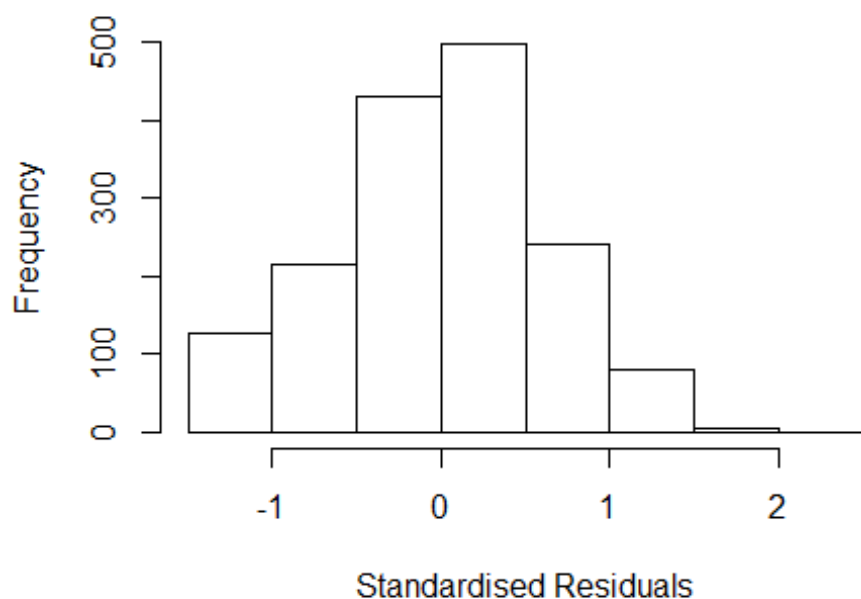
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.38781 -0.37959 -0.00134 0.37595 2.07870
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6990 0.0941 7.43 1.8e-13 ***
## FirstAuthorFemale1 0.0150 0.0440 0.34 0.73
## LastAuthorFemale1 -0.0719 0.0556 -1.29 0.20
## UniqueAuthors2 0.5050 0.0440 11.48 < 2e-16 ***
## UniqueAuthors3 0.5384 0.0471 11.45 < 2e-16 ***
## UniqueAuthors4 0.5606 0.0447 12.54 < 2e-16 ***
## UniqueAuthors5 0.6224 0.0456 13.64 < 2e-16 ***
## Year1997 -0.0018 0.1122 -0.02 0.99
## Year1998 -0.0519 0.1086 -0.48 0.63
## Year1999 -0.1225 0.1053 -1.16 0.24
```

```

## Year2000          0.0561      0.1088      0.52      0.61
## Year2001          0.0728      0.1255      0.58      0.56
## Year2002         -0.0811      0.1125     -0.72      0.47
## Year2003         -0.0764      0.1085     -0.70      0.48
## Year2004         -0.0448      0.1078     -0.42      0.68
## Year2005         -0.1044      0.1073     -0.97      0.33
## Year2006         -0.1076      0.1085     -0.99      0.32
## Year2007         -0.1066      0.1045     -1.02      0.31
## Year2008          0.0421      0.1038      0.41      0.68
## Year2009          0.0664      0.1012      0.66      0.51
## Year2010         -0.1120      0.1032     -1.09      0.28
## Year2011         -0.1220      0.1052     -1.16      0.25
## Year2012         -0.1716      0.1090     -1.57      0.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.176, Adjusted R-squared:  0.165
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 113 weights are ~= 1. The remaining 1483 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.132  0.877  0.949   0.908   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.069 1      1.034
## LastAuthorFemale  1.048 1      1.024
## Year              1.068 16      1.002

```

## Residuals from first and last author

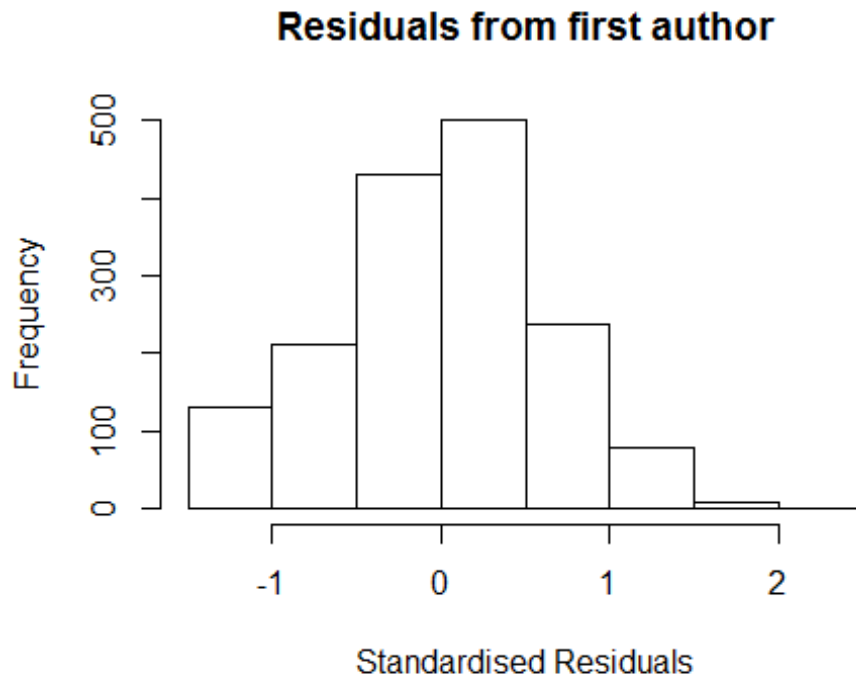


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3081 -0.4214 0.0161 0.4157 2.1096
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.04111 0.10417 9.99 <2e-16 ***
## FirstAuthorFemale1 0.09266 0.04668 1.98 0.047 *
## LastAuthorFemale1 -0.07187 0.06262 -1.15 0.251
## Year1997 -0.00462 0.13001 -0.04 0.972
## Year1998 -0.01601 0.12105 -0.13 0.895
## Year1999 -0.09361 0.12102 -0.77 0.439
## Year2000 0.04288 0.12170 0.35 0.725
## Year2001 0.07015 0.13367 0.52 0.600
## Year2002 -0.07161 0.13110 -0.55 0.585
## Year2003 0.04628 0.12044 0.38 0.701
## Year2004 0.02017 0.12242 0.16 0.869
## Year2005 -0.00169 0.12171 -0.01 0.989
```

```

## Year2006          -0.10009      0.12401    -0.81      0.420
## Year2007          -0.03064      0.11907    -0.26      0.797
## Year2008           0.11005      0.11515      0.96      0.339
## Year2009           0.17433      0.11578      1.51      0.132
## Year2010          -0.01711      0.11932    -0.14      0.886
## Year2011          -0.03573      0.12123    -0.29      0.768
## Year2012          -0.10205      0.12394    -0.82      0.410
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.616
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.00647
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.217  0.864  0.950  0.910  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.048 1      1.023
## Year              1.048 16      1.001

```



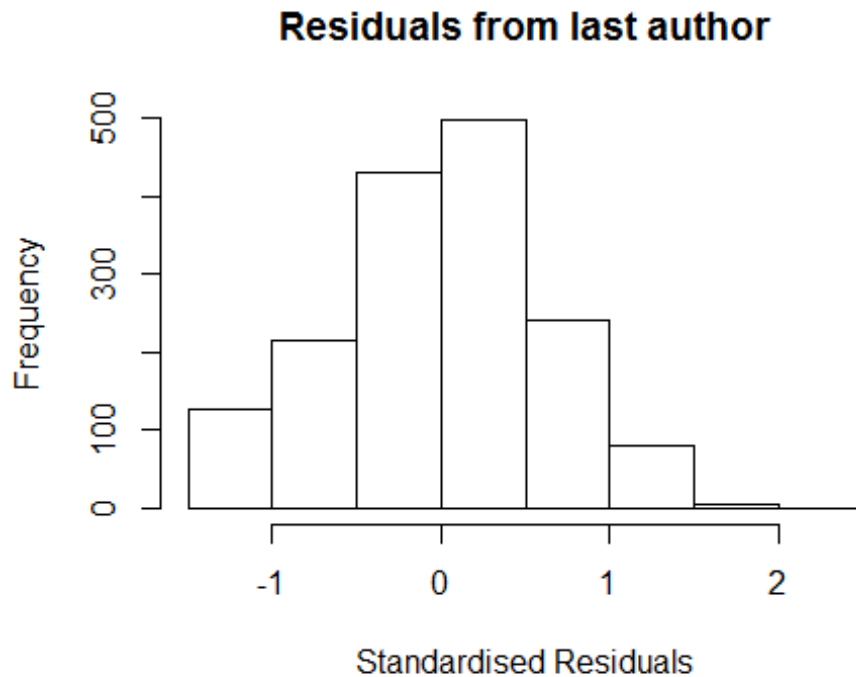
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2887 -0.4240 0.0204 0.4178 2.1144
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.03759 0.10416 9.96 <2e-16 ***
## FirstAuthorFemale1 0.07706 0.04746 1.62 0.10
## Year1997 -0.00442 0.13009 -0.03 0.97
## Year1998 -0.01427 0.12114 -0.12 0.91
## Year1999 -0.09258 0.12111 -0.76 0.44
## Year2000 0.04171 0.12181 0.34 0.73
## Year2001 0.06841 0.13340 0.51 0.61
## Year2002 -0.07227 0.13076 -0.55 0.58
## Year2003 0.04503 0.12036 0.37 0.71
## Year2004 0.02124 0.12234 0.17 0.86
## Year2005 -0.00104 0.12149 -0.01 0.99
## Year2006 -0.10136 0.12395 -0.82 0.41
```



```

## Year2007          -0.02721    0.11911   -0.23    0.82
## Year2008          0.10731    0.11533    0.93    0.35
## Year2009          0.17401    0.11582    1.50    0.13
## Year2010         -0.01963    0.11957   -0.16    0.87
## Year2011         -0.03633    0.12122   -0.30    0.76
## Year2012         -0.10532    0.12424   -0.85    0.40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.617
## Multiple R-squared:  0.0169, Adjusted R-squared:  0.00627
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 133 weights are ~= 1. The remaining 1463 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.216  0.865  0.950  0.911  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.025 1          1.012
## Year            1.025 16          1.001

```



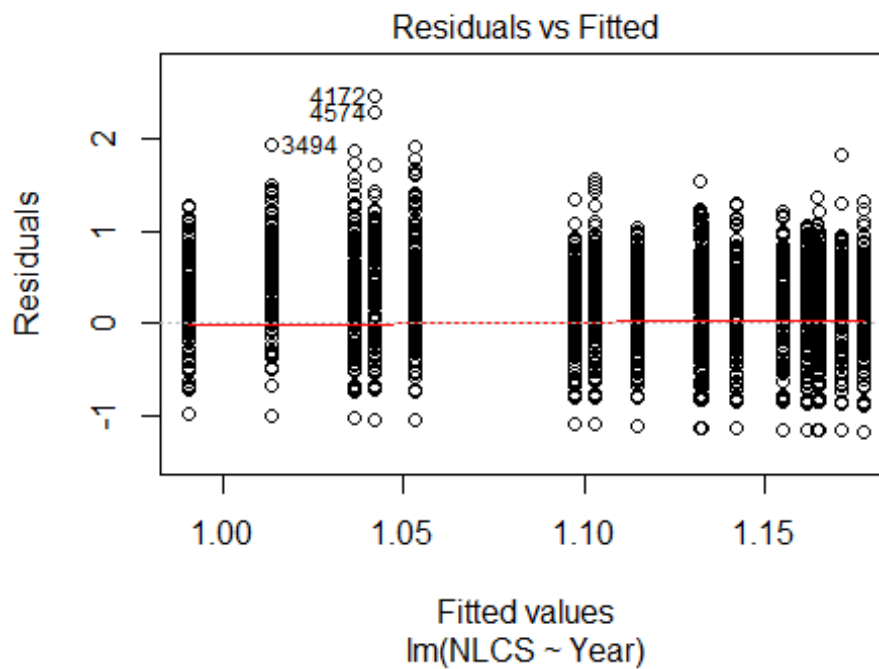
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2285 -0.4146 0.0234 0.4218 2.1019
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.05253 0.10315 10.20 <2e-16 ***
## LastAuthorFemale1 -0.04157 0.06055 -0.69 0.49
## Year1997 -0.00909 0.12926 -0.07 0.94
## Year1998 -0.01694 0.12067 -0.14 0.89
## Year1999 -0.09874 0.12037 -0.82 0.41
## Year2000 0.04521 0.12161 0.37 0.71
## Year2001 0.06255 0.13319 0.47 0.64
## Year2002 -0.07059 0.13061 -0.54 0.59
## Year2003 0.04252 0.12006 0.35 0.72
## Year2004 0.01813 0.12225 0.15 0.88
## Year2005 -0.00273 0.12131 -0.02 0.98
## Year2006 -0.09822 0.12310 -0.80 0.43
```

```

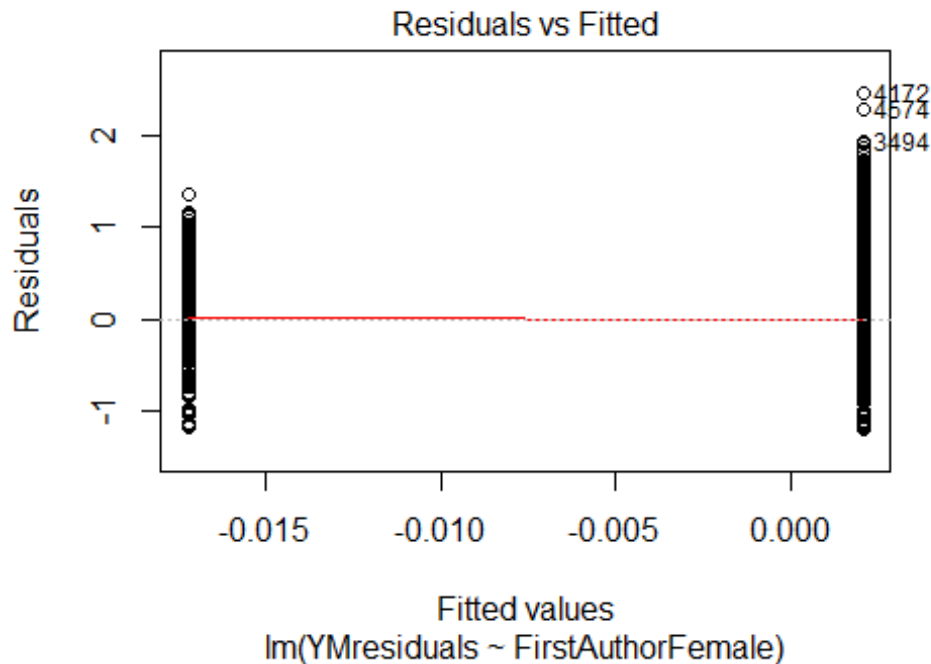
## Year2007          -0.03654      0.11837    -0.31      0.76
## Year2008           0.10907      0.11462      0.95      0.34
## Year2009           0.17598      0.11525      1.53      0.13
## Year2010          -0.01611      0.11928     -0.14      0.89
## Year2011          -0.03995      0.12074     -0.33      0.74
## Year2012          -0.11004      0.12298     -0.89      0.37
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0154, Adjusted R-squared:  0.00475
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1452 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.225  0.864  0.950  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1596"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3106"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 423 457 491 547 532 465 491 298 368 345 353 405 367 454 360
## 2011 2012
## 417 343
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 98 153 168 199 220 201 234 163 191 191 170 183 194 220 188
## 2011 2012

```

```
## 210 189
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 145 160 183 208 188 207 146 176 164 152 162 177 197 165
## 2011 2012
## 193 166
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

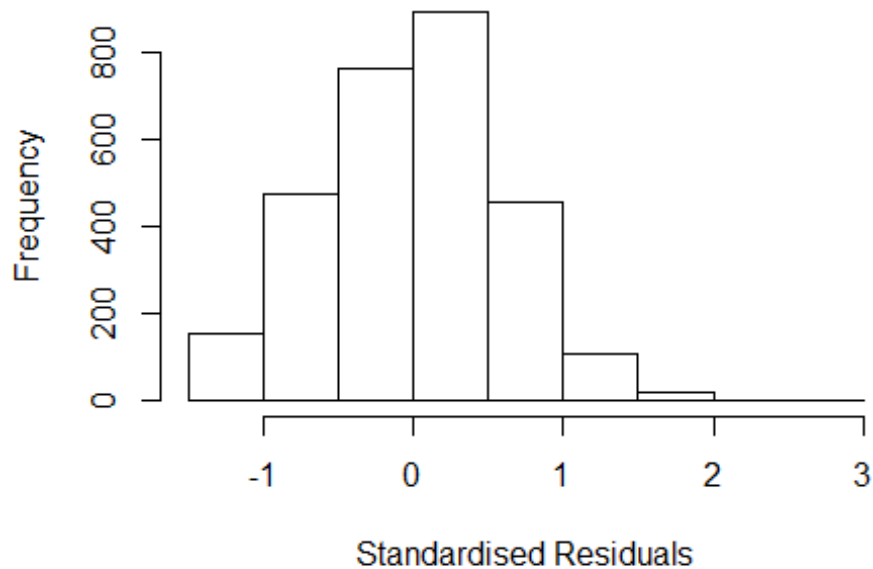


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.01, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 3.32593870684733"
## [1] "Male first author team size 2018 geometric mean: 2.10232663927312"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1200, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.06936094886239"
## [1] "Male last author team size 2018 geometric mean: 2.30796595092631"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 460, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.062 1          1.030
## LastAuthorFemale  1.051 1          1.025
## UniqueAuthors    1.122 4          1.014
## Year             1.132 16          1.004
```

## Residuals from first and last author and team size



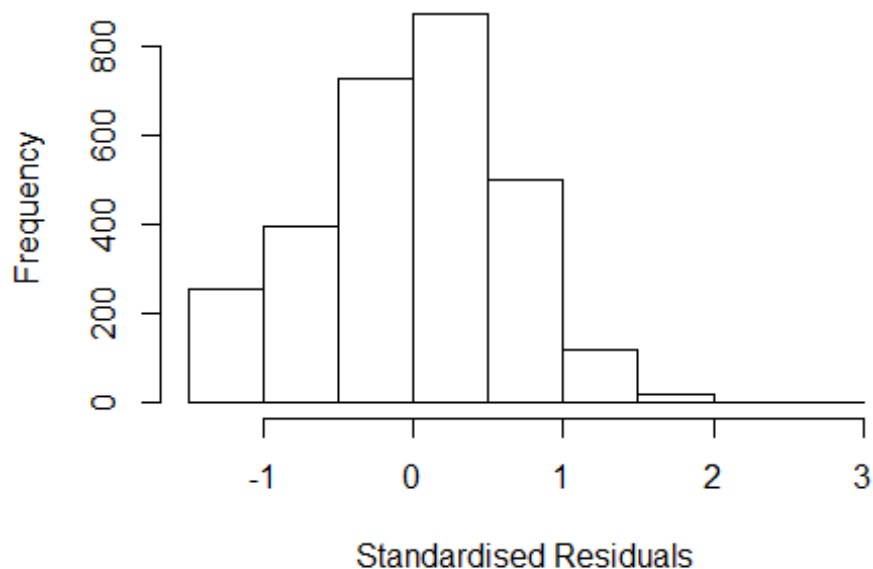
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 4172 0037113847 3.5 2002      3101      2      2.695
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4172 -0.4424  0.0276  0.4246  2.6948
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.97e-01   7.49e-02  13.31  < 2e-16 ***
## FirstAuthorFemale1 -5.88e-05   3.95e-02   0.00  0.99881
## LastAuthorFemale1 -2.99e-02   4.04e-02  -0.74  0.45862
## UniqueAuthors2    2.65e-01   2.83e-02   9.38  < 2e-16 ***
## UniqueAuthors3    3.84e-01   3.66e-02  10.47  < 2e-16 ***
## UniqueAuthors4    3.70e-01   4.93e-02   7.50  8.3e-14 ***
## UniqueAuthors5    2.28e-01   6.27e-02   3.63  0.00029 ***
## Year1997          3.51e-02   8.79e-02   0.40  0.68939
## Year1998          3.68e-02   8.72e-02   0.42  0.67294
## Year1999         -1.30e-01   9.51e-02  -1.37  0.17135
```

```

## Year2000      -1.34e-01  9.55e-02  -1.41  0.15932
## Year2001      -1.56e-01  1.01e-01  -1.54  0.12276
## Year2002      -1.92e-01  9.20e-02  -2.08  0.03750 *
## Year2003      -1.07e-01  9.24e-02  -1.16  0.24695
## Year2004      -2.03e-01  9.01e-02  -2.25  0.02442 *
## Year2005      -7.23e-02  8.58e-02  -0.84  0.39955
## Year2006      -5.73e-02  8.62e-02  -0.66  0.50640
## Year2007      -1.77e-02  8.51e-02  -0.21  0.83510
## Year2008      -1.29e-01  8.26e-02  -1.56  0.11881
## Year2009      -5.21e-02  8.30e-02  -0.63  0.53053
## Year2010      -4.85e-02  8.44e-02  -0.57  0.56605
## Year2011      -2.79e-02  8.41e-02  -0.33  0.74049
## Year2012      -4.88e-02  8.40e-02  -0.58  0.56143
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.619
## Multiple R-squared:  0.0685, Adjusted R-squared:  0.0614
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 233 weights are ~= 1. The remaining 2645 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.019  0.861  0.948  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.47e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.054 1          1.026
## LastAuthorFemale  1.049 1          1.024
## Year              1.028 16          1.001

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 4172 0037113847 3.5 2002    3101      2      2.506
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.180 -0.440  0.035  0.448  2.506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13124    0.07711   14.67  <2e-16 ***
## FirstAuthorFemale1  0.01283    0.04113    0.31    0.76
## LastAuthorFemale1 -0.03224    0.04319   -0.75    0.46
## Year1997         0.04850    0.09074    0.53    0.59
## Year1998         0.03633    0.09055    0.40    0.69
## Year1999        -0.09570    0.09719   -0.98    0.32
## Year2000        -0.10846    0.09877   -1.10    0.27
## Year2001        -0.11673    0.10478   -1.11    0.27
## Year2002        -0.13710    0.09623   -1.42    0.15
## Year2003        -0.06398    0.09604   -0.67    0.51
## Year2004        -0.13692    0.09386   -1.46    0.14
## Year2005        -0.01564    0.09008   -0.17    0.86
```

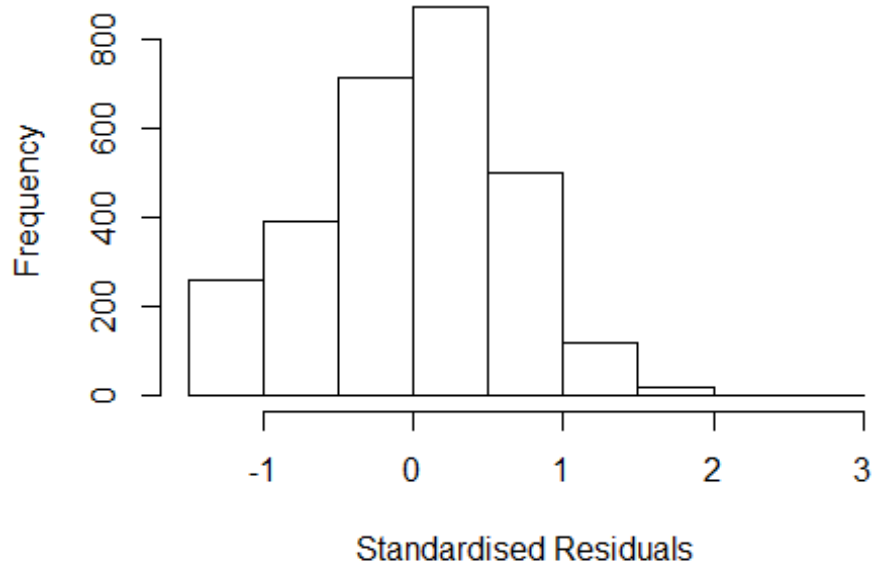


```

## Year2006      -0.00594    0.09083   -0.07    0.95
## Year2007      0.04192    0.08952    0.47    0.64
## Year2008     -0.04893    0.08555   -0.57    0.57
## Year2009      0.01217    0.08653    0.14    0.89
## Year2010      0.04416    0.08864    0.50    0.62
## Year2011      0.03358    0.08704    0.39    0.70
## Year2012      0.04217    0.08745    0.48    0.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.643
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00516
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 258 weights are ~= 1. The remaining 2620 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0951 0.8610 0.9470 0.9100 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1      1.009
## Year      1.018 16      1.001

```

## Residuals from first author



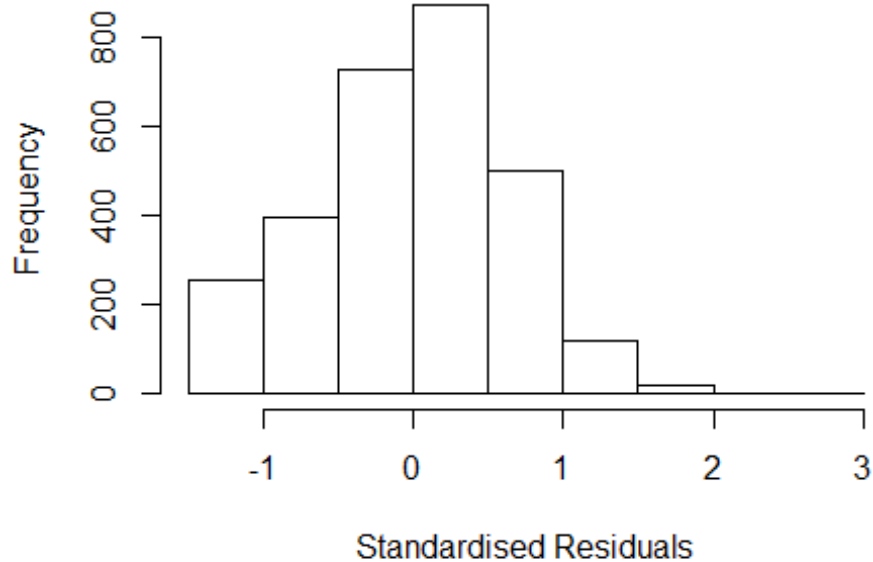
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 4172 0037113847 3.5 2002    3101      2    2.506
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1809 -0.4399  0.0344  0.4480  2.5080
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.12924    0.07702   14.66  <2e-16 ***
## FirstAuthorFemale1 0.00294    0.04108    0.07    0.94
## Year1997        0.04874    0.09075    0.54    0.59
## Year1998        0.03598    0.09058    0.40    0.69
## Year1999       -0.09571    0.09717   -0.98    0.32
## Year2000       -0.10892    0.09861   -1.10    0.27
## Year2001       -0.11662    0.10451   -1.12    0.26
## Year2002       -0.13723    0.09622   -1.43    0.15
## Year2003       -0.06375    0.09608   -0.66    0.51
## Year2004       -0.13602    0.09389   -1.45    0.15
## Year2005       -0.01519    0.09005   -0.17    0.87
## Year2006       -0.00518    0.09076   -0.06    0.95
```

```

## Year2007          0.04177    0.08956    0.47    0.64
## Year2008         -0.04857    0.08557   -0.57    0.57
## Year2009          0.01195    0.08656    0.14    0.89
## Year2010          0.04530    0.08864    0.51    0.61
## Year2011          0.03425    0.08705    0.39    0.69
## Year2012          0.04120    0.08747    0.47    0.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.643
## Multiple R-squared:  0.0112, Adjusted R-squared:  0.00535
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 256 weights are ~= 1. The remaining 2622 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0935 0.8610 0.9470 0.9100 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.012 1          1.006
## Year            1.012 16          1.000

```

## Residuals from last author



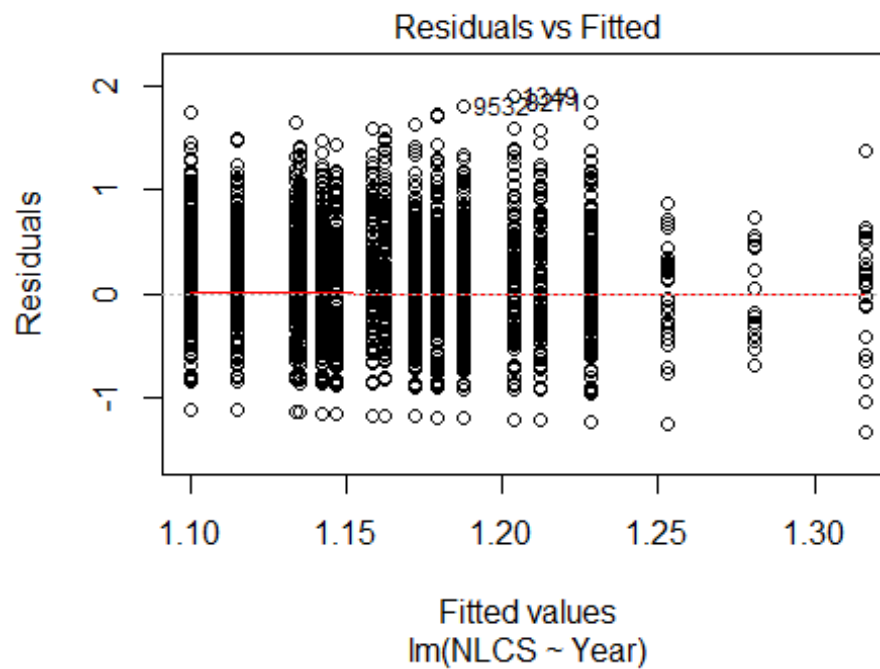
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 4172 0037113847 3.5 2002    3101    2    2.506
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1809 -0.4415  0.0345  0.4474  2.5047
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13234    0.07711   14.68  <2e-16 ***
## LastAuthorFemale1 -0.02749    0.04282   -0.64    0.52
## Year1997         0.04859    0.09085    0.53    0.59
## Year1998         0.03529    0.09056    0.39    0.70
## Year1999        -0.09603    0.09733   -0.99    0.32
## Year2000        -0.10876    0.09880   -1.10    0.27
## Year2001        -0.11684    0.10503   -1.11    0.27
## Year2002        -0.13708    0.09636   -1.42    0.15
## Year2003        -0.06447    0.09619   -0.67    0.50
## Year2004        -0.13687    0.09404   -1.46    0.15
## Year2005        -0.01538    0.09018   -0.17    0.86
## Year2006        -0.00565    0.09094   -0.06    0.95
```

```

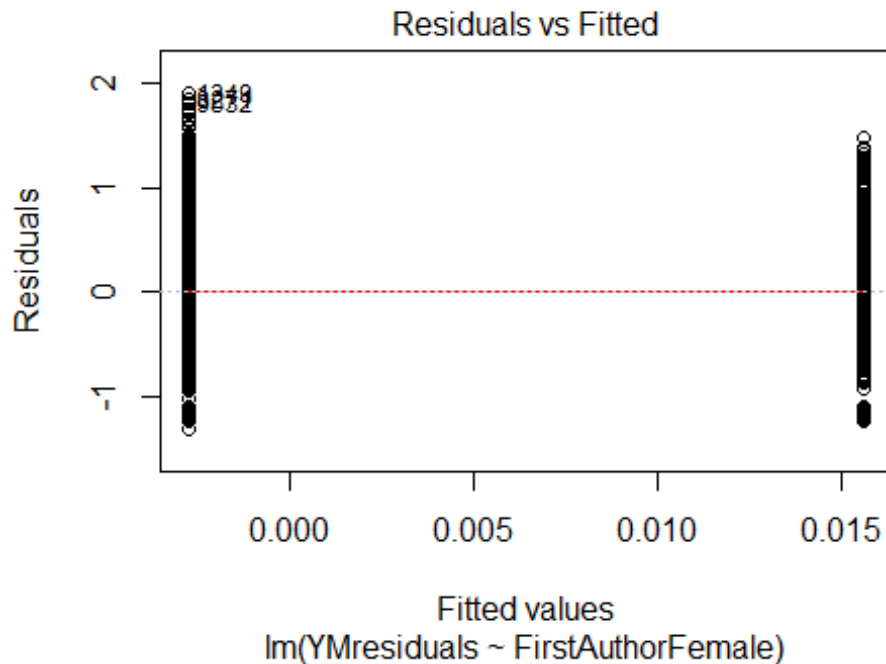
## Year2007      0.04184      0.08961      0.47      0.64
## Year2008     -0.04903      0.08566     -0.57      0.57
## Year2009      0.01223      0.08662      0.14      0.89
## Year2010      0.04417      0.08874      0.50      0.62
## Year2011      0.03366      0.08713      0.39      0.70
## Year2012      0.04206      0.08753      0.48      0.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.64
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00553
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 255 weights are ~= 1. The remaining 2623 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0908 0.8590 0.9460 0.9090 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.47e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2878"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3107"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 637 513 534 89 795 625 592 571 42 590 87 644 650 677 666
## 2011 2012
## 669 638
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 130 129 129 15 213 163 212 225 21 221 35 259 248 303 330
## 2011 2012

```

```
## 314 311
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 109 117 14 180 141 166 189 18 190 33 236 222 259 283
## 2011 2012
## 269 266
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 63, df = 16, p-value = 1e-07
```

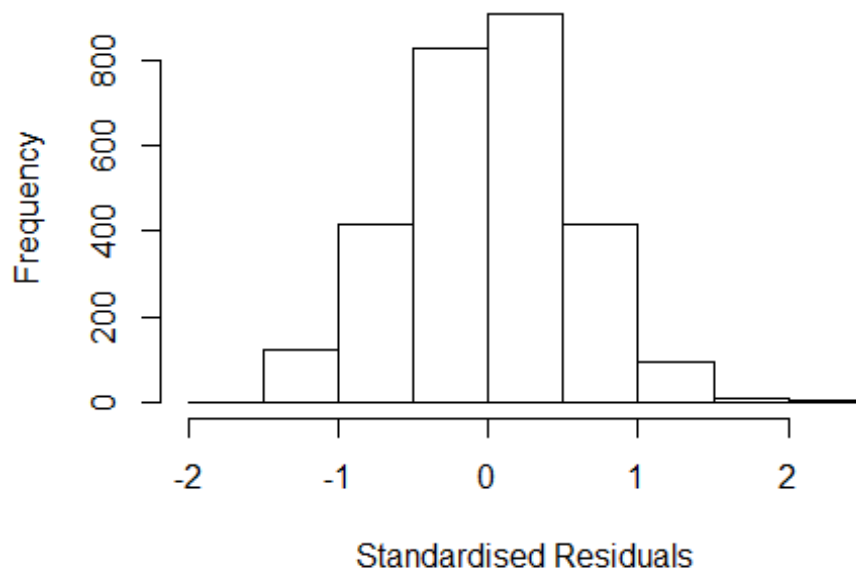


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.4, df = 1, p-value = 0.06
```



```
## [1] "Female first author team size 2018 geometric mean: 3.75244430250331"
## [1] "Male first author team size 2018 geometric mean: 3.44209372223121"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8600, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.53777305431298"
## [1] "Male last author team size 2018 geometric mean: 3.48243695001976"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6200, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.083 1          1.041
## LastAuthorFemale  1.050 1          1.025
## UniqueAuthors     1.181 4          1.021
## Year              1.201 16         1.006
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5847 -0.3841 0.0146 0.3859 2.1184
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86781 0.06726 12.90 < 2e-16 ***
## FirstAuthorFemale1 -0.03601 0.03066 -1.17 0.240
## LastAuthorFemale1 0.02915 0.04165 0.70 0.484
## UniqueAuthors2 0.26464 0.03719 7.12 1.4e-12 ***
## UniqueAuthors3 0.31584 0.03706 8.52 < 2e-16 ***
## UniqueAuthors4 0.39594 0.04115 9.62 < 2e-16 ***
## UniqueAuthors5 0.50926 0.03802 13.39 < 2e-16 ***
## Year1997 0.09562 0.09765 0.98 0.328
## Year1998 0.07644 0.09074 0.84 0.400
## Year1999 0.01062 0.15359 0.07 0.945
```

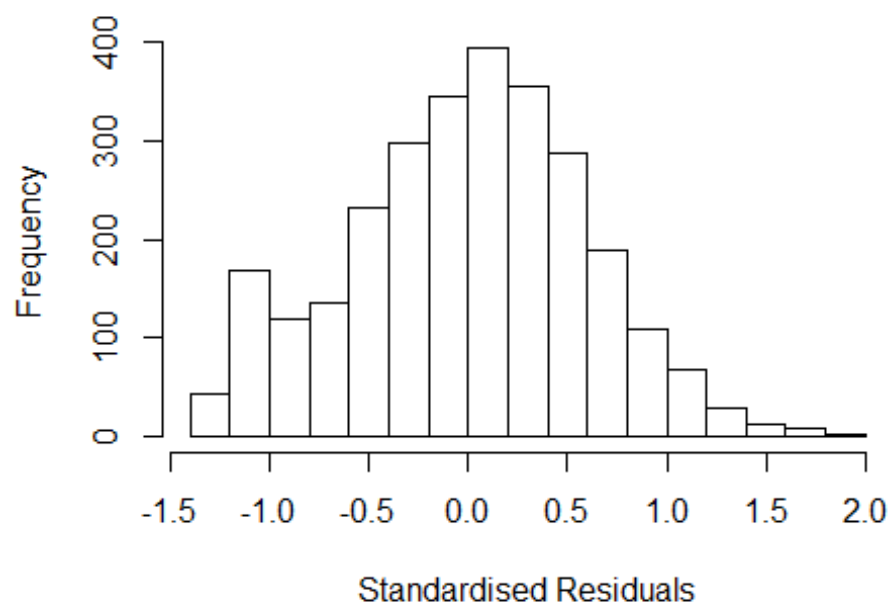


```

## Year2000          0.03026    0.08492    0.36    0.722
## Year2001          0.04747    0.08867    0.54    0.592
## Year2002          0.05188    0.08069    0.64    0.520
## Year2003         -0.00575    0.08176   -0.07    0.944
## Year2004          0.22251    0.19341    1.15    0.250
## Year2005         -0.01763    0.07519   -0.23    0.815
## Year2006          0.20767    0.12024    1.73    0.084 .
## Year2007          0.00320    0.07241    0.04    0.965
## Year2008          0.07877    0.07399    1.06    0.287
## Year2009          0.02261    0.07241    0.31    0.755
## Year2010          0.03693    0.07144    0.52    0.605
## Year2011         -0.03759    0.07240   -0.52    0.604
## Year2012         -0.05098    0.07148   -0.71    0.476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.0782, Adjusted R-squared:  0.0709
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 214 weights are ~= 1. The remaining 2586 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.139  0.866  0.952  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.056 1      1.028
## LastAuthorFemale  1.044 1      1.022
## Year              1.057 16      1.002

```

## Residuals from first and last author



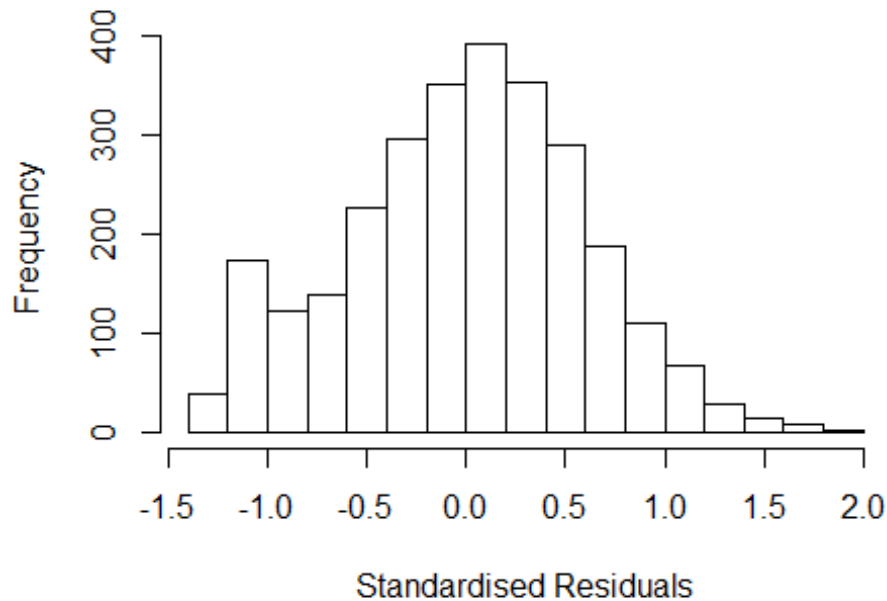
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3429 -0.3989  0.0309  0.4048  1.9460
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0832     0.0668   16.21  <2e-16 ***
## FirstAuthorFemale1  0.0172     0.0311    0.55   0.580
## LastAuthorFemale1  0.0331     0.0431    0.77   0.442
## Year1997           0.0749     0.0991    0.76   0.450
## Year1998           0.1191     0.0962    1.24   0.216
## Year1999           0.1486     0.1433    1.04   0.300
## Year2000           0.0484     0.0901    0.54   0.591
## Year2001           0.0714     0.0910    0.79   0.432
## Year2002           0.0845     0.0839    1.01   0.314
## Year2003           0.0525     0.0848    0.62   0.536
## Year2004           0.2598     0.1772    1.47   0.143
## Year2005           0.0490     0.0790    0.62   0.535
```

```

## Year2006          0.2050      0.1083      1.89      0.058 .
## Year2007          0.0834      0.0756      1.10      0.270
## Year2008          0.1430      0.0776      1.84      0.065 .
## Year2009          0.0699      0.0761      0.92      0.359
## Year2010          0.1348      0.0748      1.80      0.072 .
## Year2011          0.0525      0.0753      0.70      0.485
## Year2012          0.0601      0.0749      0.80      0.422
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.00488,    Adjusted R-squared:  -0.00156
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 230 weights are ~= 1. The remaining 2570 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.254  0.865  0.949  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.034 1      1.017
## Year              1.034 16      1.001

```

## Residuals from first author



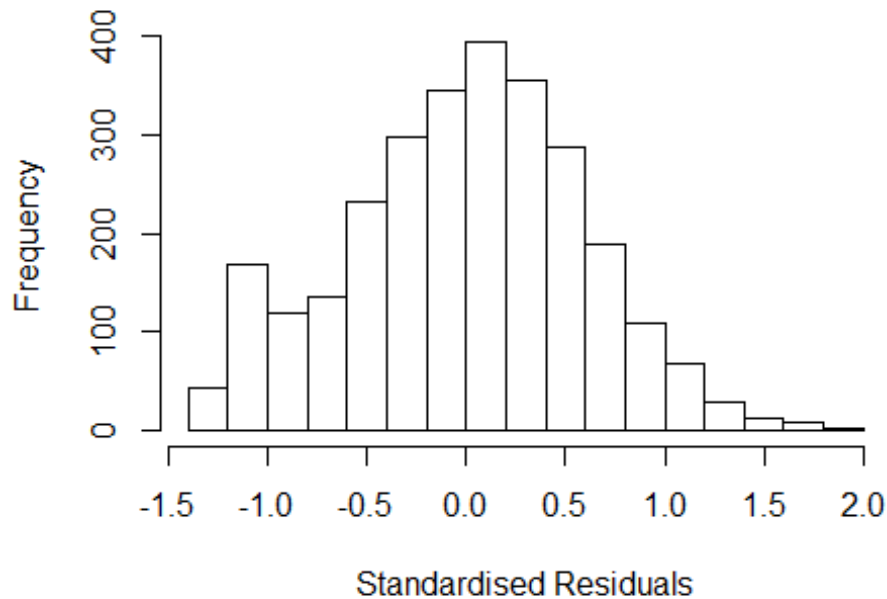
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3440 -0.3996  0.0288  0.4063  1.9437
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0853     0.0667   16.28  <2e-16 ***
## FirstAuthorFemale1  0.0218     0.0308    0.71   0.480
## Year1997          0.0750     0.0991    0.76   0.449
## Year1998          0.1175     0.0961    1.22   0.221
## Year1999          0.1511     0.1432    1.06   0.291
## Year2000          0.0478     0.0902    0.53   0.596
## Year2001          0.0718     0.0909    0.79   0.429
## Year2002          0.0851     0.0839    1.01   0.310
## Year2003          0.0523     0.0849    0.62   0.538
## Year2004          0.2587     0.1771    1.46   0.144
## Year2005          0.0477     0.0789    0.60   0.546
## Year2006          0.2037     0.1081    1.88   0.060 .
```

```

## Year2007          0.0832      0.0755      1.10      0.271
## Year2008          0.1425      0.0775      1.84      0.066 .
## Year2009          0.0699      0.0761      0.92      0.358
## Year2010          0.1336      0.0747      1.79      0.074 .
## Year2011          0.0518      0.0752      0.69      0.491
## Year2012          0.0614      0.0748      0.82      0.412
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.589
## Multiple R-squared:  0.00468,    Adjusted R-squared:  -0.00141
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 234 weights are ~= 1. The remaining 2566 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.254  0.865  0.948  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.011
## Year            1.023 16          1.001

```

## Residuals from last author



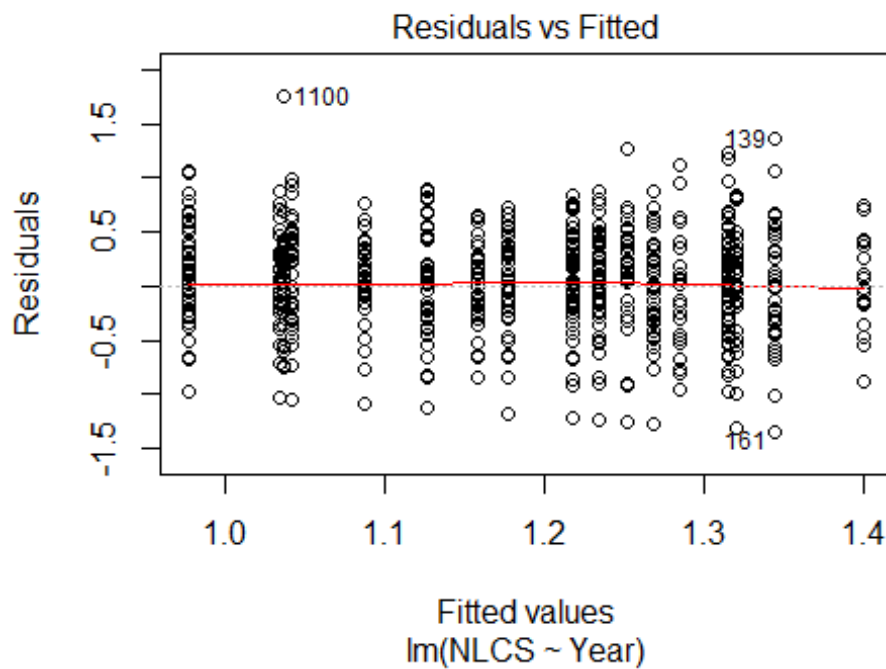
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3486 -0.4006 0.0292 0.4053 1.9448
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0844 0.0668 16.23 <2e-16 ***
## LastAuthorFemale1 0.0373 0.0426 0.88 0.381
## Year1997 0.0748 0.0990 0.76 0.450
## Year1998 0.1194 0.0962 1.24 0.215
## Year1999 0.1480 0.1431 1.03 0.301
## Year2000 0.0497 0.0901 0.55 0.581
## Year2001 0.0711 0.0910 0.78 0.435
## Year2002 0.0855 0.0839 1.02 0.308
## Year2003 0.0543 0.0848 0.64 0.522
## Year2004 0.2641 0.1767 1.50 0.135
## Year2005 0.0514 0.0788 0.65 0.514
## Year2006 0.2048 0.1084 1.89 0.059 .
```

```

## Year2007          0.0839      0.0756      1.11      0.267
## Year2008          0.1448      0.0775      1.87      0.062 .
## Year2009          0.0704      0.0761      0.92      0.355
## Year2010          0.1363      0.0747      1.82      0.068 .
## Year2011          0.0536      0.0753      0.71      0.476
## Year2012          0.0611      0.0749      0.82      0.414
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.59
## Multiple R-squared:  0.00479,    Adjusted R-squared:  -0.0013
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 229 weights are ~= 1. The remaining 2571 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.254  0.866  0.949  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.57e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2800"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3108"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   99  137   91  138  120  120   63   81   75   77   76   76   80  104   80
## 2011 2012
##   78  109
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   27   40   32   40   40   35   18   38   35   42   36   54   54   59   39
## 2011 2012

```

```
## 34 53
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 21 37 30 34 35 29 16 30 27 32 25 35 41 53 32
## 2011 2012
## 30 44
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.6
```

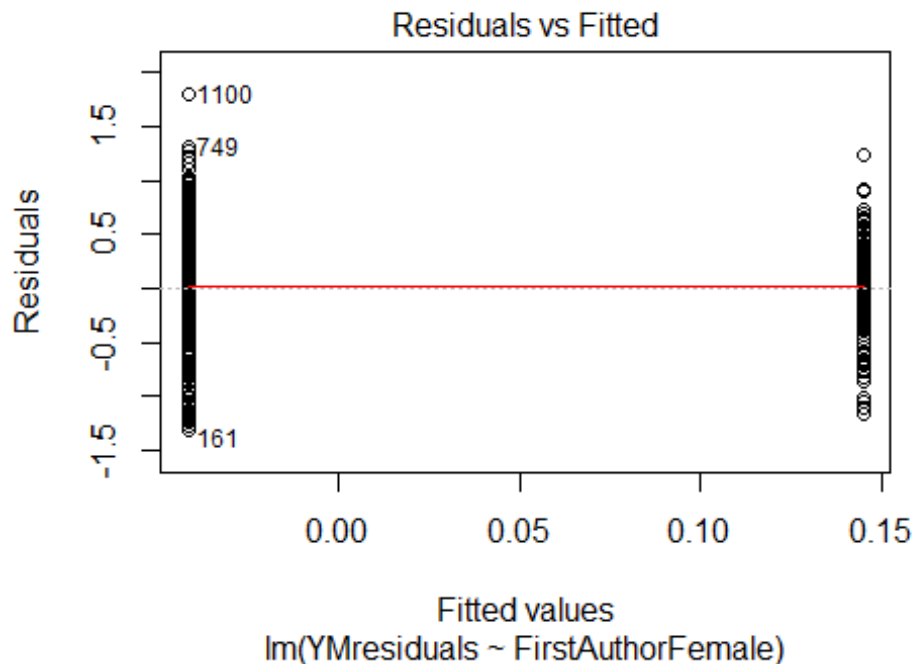


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 4e-04
## [1] "Female first author team size 2018 geometric mean: 5.5104203843964"
## [1] "Male first author team size 2018 geometric mean: 3.58039205162856"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



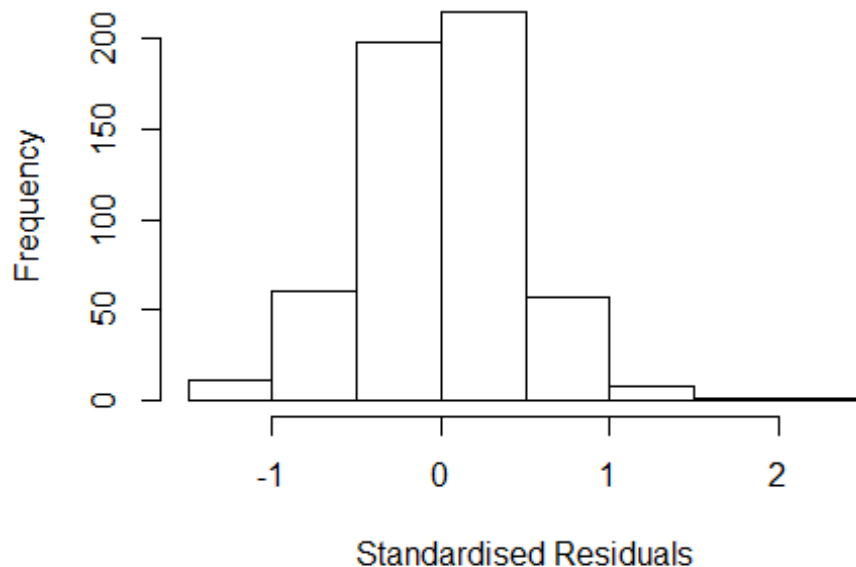
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 7.57326879575983"
## [1] "Male last author team size 2018 geometric mean: 3.57111575213485"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.216 1      1.103
## LastAuthorFemale  1.301 1      1.141
## UniqueAuthors    2.464 4      1.119
## Year              2.522 16     1.029
```

## Residuals from first and last author and team size



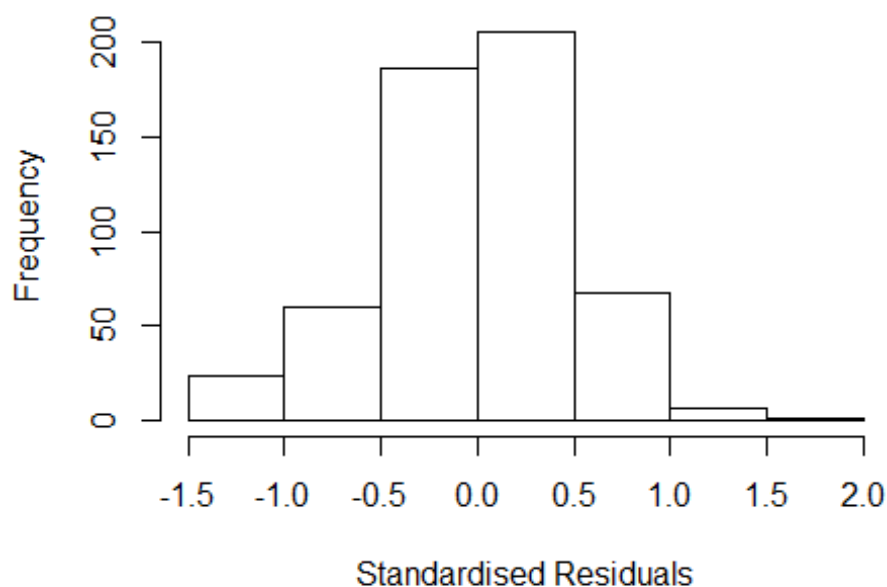
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3492 -0.2813 0.0152 0.2887 2.0489
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90897 0.11931 7.62 1.2e-13 ***
## FirstAuthorFemale1 0.10076 0.04238 2.38 0.01778 *
## LastAuthorFemale1 0.06803 0.05958 1.14 0.25406
## UniqueAuthors2 0.16302 0.07778 2.10 0.03658 *
## UniqueAuthors3 0.20153 0.07983 2.52 0.01188 *
## UniqueAuthors4 0.27995 0.08308 3.37 0.00081 ***
## UniqueAuthors5 0.46180 0.06984 6.61 9.3e-11 ***
## Year1997 0.23489 0.14481 1.62 0.10538
## Year1998 0.00727 0.12385 0.06 0.95321
## Year1999 -0.10097 0.15025 -0.67 0.50188
```

```

## Year2000      0.13436      0.12519      1.07      0.28366
## Year2001      0.16028      0.15243      1.05      0.29352
## Year2002      0.21277      0.15595      1.36      0.17303
## Year2003     -0.16484      0.11415     -1.44      0.14930
## Year2004     -0.01324      0.14158     -0.09      0.92555
## Year2005      0.02587      0.15148      0.17      0.86446
## Year2006     -0.07966      0.12548     -0.63      0.52579
## Year2007     -0.07166      0.12951     -0.55      0.58029
## Year2008      0.07831      0.13467      0.58      0.56118
## Year2009      0.04605      0.12176      0.38      0.70546
## Year2010     -0.02377      0.12953     -0.18      0.85448
## Year2011      0.01656      0.12807      0.13      0.89713
## Year2012     -0.02731      0.12530     -0.22      0.82755
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.422
## Multiple R-squared:  0.182, Adjusted R-squared:  0.148
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 229 is an outlier with |weight| = 0 ( < 0.00018);
## 54 weights are ~= 1. The remaining 496 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.154  0.853  0.950  0.891  0.986  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.81e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1           1.044
## LastAuthorFemale  1.159 1           1.077
## Year              1.234 16           1.007

```

## Residuals from first and last author



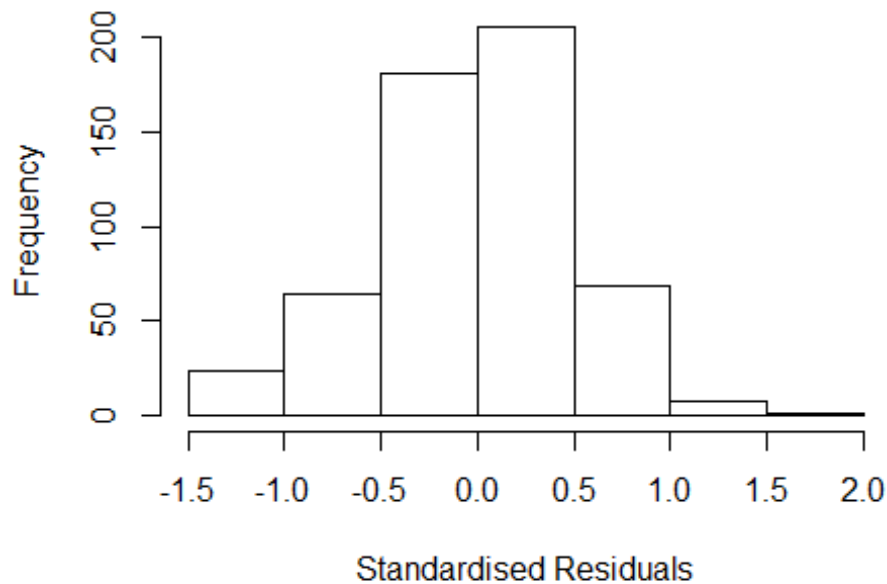
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3260 -0.3011  0.0257  0.3115  1.8093
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.06211    0.11546   9.20 < 2e-16 ***
## FirstAuthorFemale1 0.18664    0.04171   4.47 9.4e-06 ***
## LastAuthorFemale1 0.11788    0.05892   2.00 0.046 *
## Year1997         0.26390    0.14904   1.77 0.077 .
## Year1998         0.04046    0.13640   0.30 0.767
## Year1999        -0.00897    0.16641  -0.05 0.957
## Year2000         0.18192    0.13777   1.32 0.187
## Year2001         0.17142    0.15832   1.08 0.279
## Year2002         0.29415    0.15994   1.84 0.066 .
## Year2003        -0.07839    0.13320  -0.59 0.556
## Year2004         0.02267    0.14158   0.16 0.873
## Year2005         0.04386    0.15006   0.29 0.770
```

```

## Year2006      -0.06262    0.14060   -0.45    0.656
## Year2007      0.00570    0.14894    0.04    0.969
## Year2008      0.15085    0.13595    1.11    0.268
## Year2009      0.14951    0.13372    1.12    0.264
## Year2010      0.11515    0.14395    0.80    0.424
## Year2011      0.16402    0.14064    1.17    0.244
## Year2012      0.11816    0.13788    0.86    0.392
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.0766, Adjusted R-squared:  0.0454
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 503 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.101  0.870   0.955   0.900   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.094 1      1.046
## Year      1.094 16      1.003

```

## Residuals from first author



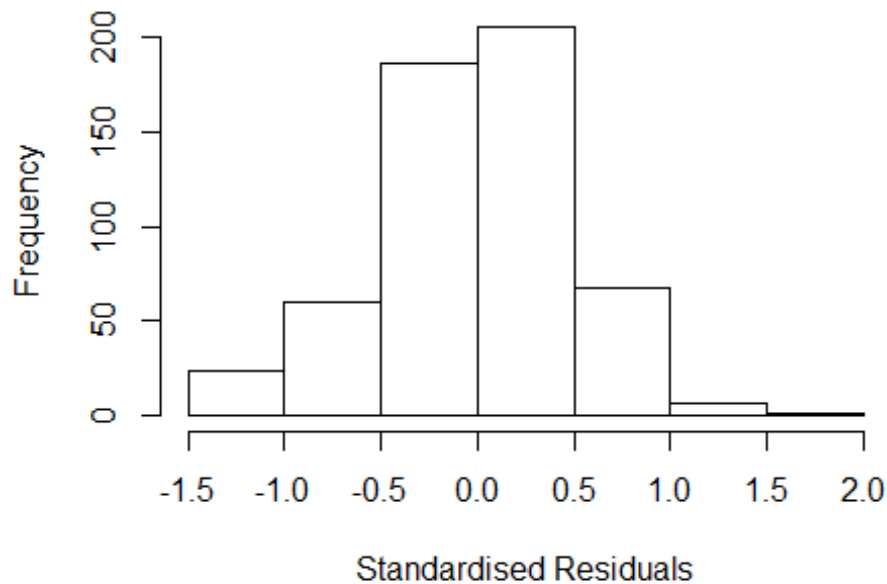
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3273 -0.3038 0.0178 0.3060 1.8018
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0868 0.1156 9.40 < 2e-16 ***
## FirstAuthorFemale1 0.2074 0.0420 4.94 1.1e-06 ***
## Year1997 0.2404 0.1477 1.63 0.104
## Year1998 0.0190 0.1375 0.14 0.890
## Year1999 -0.0235 0.1676 -0.14 0.889
## Year2000 0.1705 0.1378 1.24 0.217
## Year2001 0.1674 0.1569 1.07 0.286
## Year2002 0.2994 0.1656 1.81 0.071 .
## Year2003 -0.0956 0.1336 -0.72 0.475
## Year2004 0.0255 0.1410 0.18 0.857
## Year2005 0.0280 0.1529 0.18 0.855
## Year2006 -0.0849 0.1408 -0.60 0.547
```

```

## Year2007          -0.0171      0.1495   -0.11    0.909
## Year2008          0.1396      0.1369    1.02    0.308
## Year2009          0.1433      0.1345    1.07    0.287
## Year2010          0.0970      0.1432    0.68    0.498
## Year2011          0.1471      0.1413    1.04    0.298
## Year2012          0.1044      0.1381    0.76    0.450
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.468
## Multiple R-squared:  0.0693, Adjusted R-squared:  0.0396
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 505 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.861  0.957   0.900  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.189 1      1.090
## Year              1.189 16      1.005

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.368 -0.323  0.023  0.332  1.768
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0870     0.1196   9.09  <2e-16 ***
## LastAuthorFemale1  0.1709     0.0601   2.85  0.0046 **
## Year1997          0.2811     0.1528   1.84  0.0664 .
## Year1998          0.0477     0.1415   0.34  0.7363
## Year1999          0.0164     0.1650   0.10  0.9211
## Year2000          0.1745     0.1409   1.24  0.2161
## Year2001          0.1915     0.1615   1.19  0.2362
## Year2002          0.2827     0.1618   1.75  0.0811 .
## Year2003         -0.0618     0.1391  -0.44  0.6571
## Year2004          0.0381     0.1436   0.27  0.7908
## Year2005          0.0628     0.1530   0.41  0.6816
## Year2006         -0.0235     0.1454  -0.16  0.8718
```

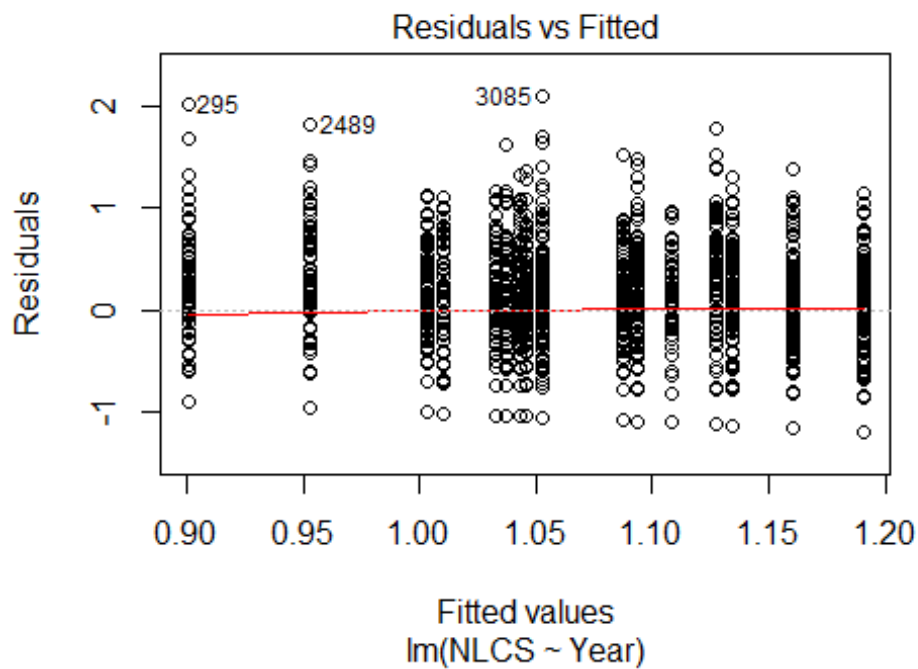


```

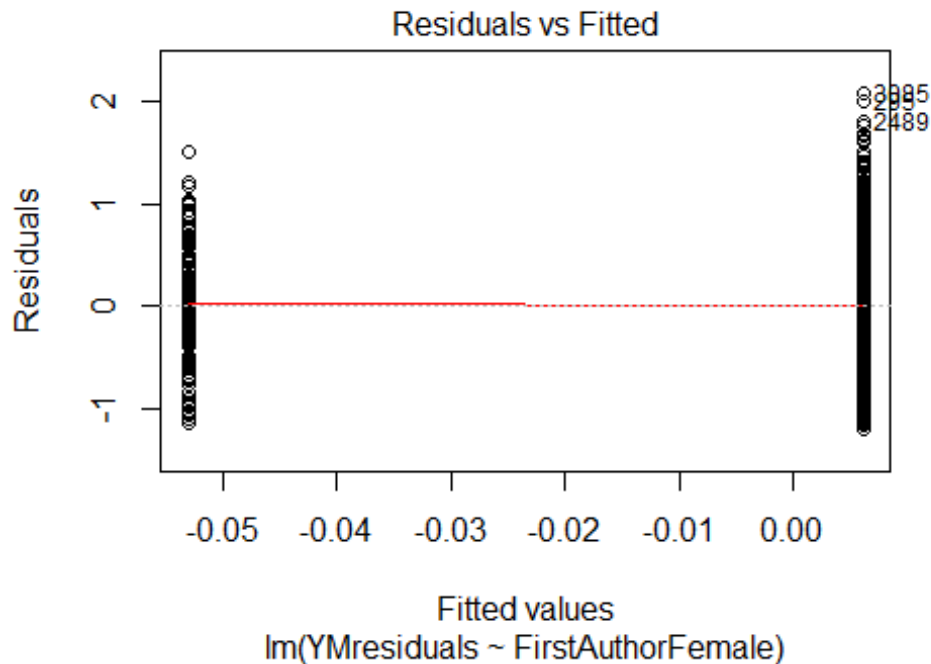
## Year2007          0.0349      0.1567      0.22      0.8240
## Year2008          0.1564      0.1426      1.10      0.2734
## Year2009          0.1590      0.1385      1.15      0.2514
## Year2010          0.1462      0.1492      0.98      0.3279
## Year2011          0.1877      0.1443      1.30      0.1938
## Year2012          0.1404      0.1406      1.00      0.3184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.478
## Multiple R-squared:  0.0509, Adjusted R-squared:  0.0206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 513 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.141  0.876  0.951  0.903  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.81e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 551"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3109"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  243  222  239  234  257  306  273  331  269  260  260  273  274  286  217
## 2011 2012
##  232  241
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   92   66   87   63   84   96  106  120  107  111  100  123  124  137  114
## 2011 2012

```

```
## 118 136
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 81 60 80 59 75 85 95 105 100 99 94 113 109 122 99
## 2011 2012
## 105 123
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 64, df = 16, p-value = 1e-07
```

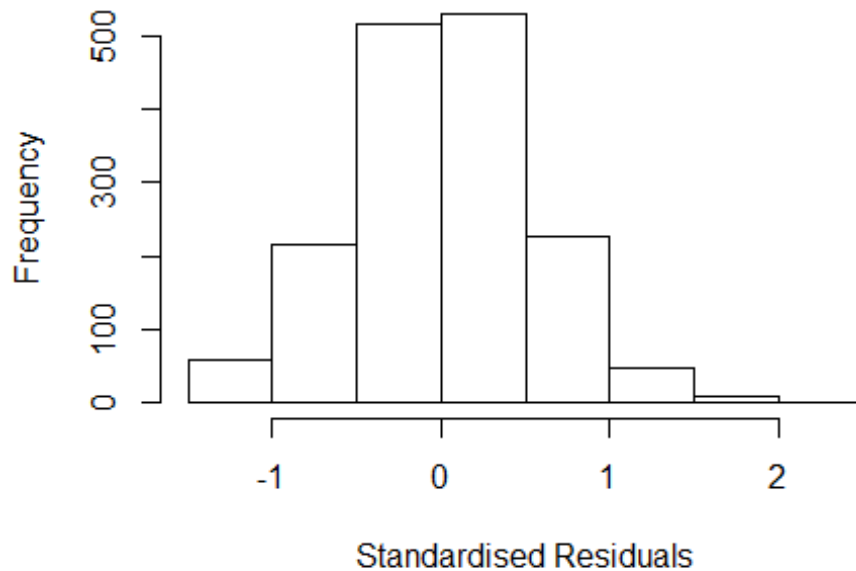


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.25, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 2.02153015930144"
## [1] "Male first author team size 2018 geometric mean: 1.88399924657436"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 730, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.40523492075426"
## [1] "Male last author team size 2018 geometric mean: 1.85883404409059"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 810, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.269 1      1.127
## LastAuthorFemale  1.292 1      1.137
## UniqueAuthors    1.258 4      1.029
## Year             1.343 16      1.009
```

## Residuals from first and last author and team size



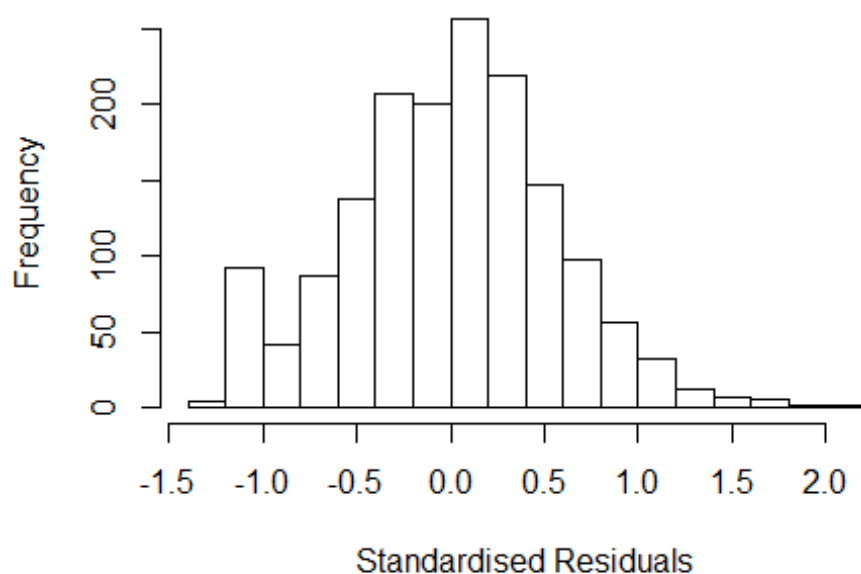
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3676 -0.3537 0.0111 0.3559 2.0631
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.76378 0.07257 10.53 < 2e-16 ***
## FirstAuthorFemale1 -0.02958 0.04898 -0.60 0.54597
## LastAuthorFemale1 -0.00536 0.06275 -0.09 0.93201
## UniqueAuthors2 0.16211 0.03134 5.17 2.6e-07 ***
## UniqueAuthors3 0.31505 0.04692 6.72 2.6e-11 ***
## UniqueAuthors4 0.39832 0.07528 5.29 1.4e-07 ***
## UniqueAuthors5 0.28318 0.08004 3.54 0.00042 ***
## Year1997 0.13308 0.10792 1.23 0.21772
## Year1998 0.13803 0.09285 1.49 0.13731
## Year1999 0.23952 0.09740 2.46 0.01404 *
```

```

## Year2000          0.18015      0.09047      1.99  0.04662 *
## Year2001          0.28877      0.11678      2.47  0.01351 *
## Year2002          0.23669      0.08662      2.73  0.00636 **
## Year2003          0.03504      0.10291      0.34  0.73354
## Year2004          0.15406      0.09718      1.59  0.11308
## Year2005          0.14611      0.09084      1.61  0.10793
## Year2006          0.17214      0.08841      1.95  0.05171 .
## Year2007          0.12767      0.08441      1.51  0.13060
## Year2008          0.15763      0.08649      1.82  0.06856 .
## Year2009          0.21945      0.08511      2.58  0.01001 *
## Year2010          0.31062      0.08829      3.52  0.00045 ***
## Year2011          0.26859      0.08618      3.12  0.00186 **
## Year2012          0.21238      0.08490      2.50  0.01247 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0726, Adjusted R-squared:  0.0597
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 146 weights are ~= 1. The remaining 1458 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.103  0.867  0.950  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.251 1      1.119
## LastAuthorFemale  1.281 1      1.132
## Year              1.120 16      1.004

```

## Residuals from first and last author



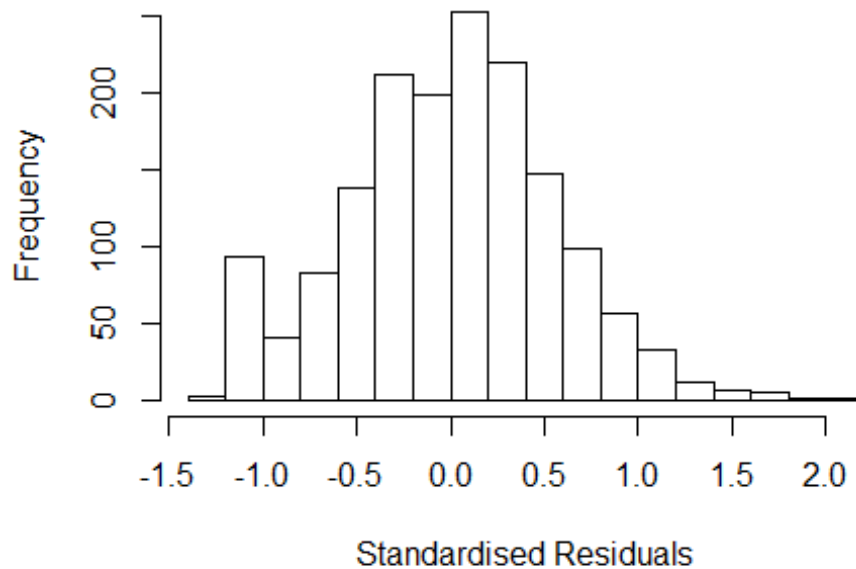
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2076 -0.3567 0.0302 0.3639 2.1126
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.84792 0.07198 11.78 < 2e-16 ***
## FirstAuthorFemale1 0.00861 0.04958 0.17 0.86208
## LastAuthorFemale1 -0.03638 0.06448 -0.56 0.57274
## Year1997 0.16361 0.11327 1.44 0.14883
## Year1998 0.17973 0.09587 1.87 0.06101 .
## Year1999 0.26196 0.09859 2.66 0.00796 **
## Year2000 0.20108 0.09336 2.15 0.03140 *
## Year2001 0.35108 0.11775 2.98 0.00291 **
## Year2002 0.25637 0.08823 2.91 0.00371 **
## Year2003 0.07854 0.10339 0.76 0.44754
## Year2004 0.18248 0.09886 1.85 0.06510 .
## Year2005 0.17707 0.09230 1.92 0.05524 .
```

```

## Year2006          0.18220    0.08898    2.05  0.04076 *
## Year2007          0.15606    0.08501    1.84  0.06658 .
## Year2008          0.18947    0.08800    2.15  0.03146 *
## Year2009          0.27538    0.08758    3.14  0.00170 **
## Year2010          0.35592    0.08914    3.99  6.8e-05 ***
## Year2011          0.31479    0.08804    3.58  0.00036 ***
## Year2012          0.28225    0.08692    3.25  0.00119 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.025, Adjusted R-squared:  0.0139
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 125 weights are ~= 1. The remaining 1479 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0976 0.8680 0.9530 0.9040 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.052 1      1.026
## Year              1.052 16      1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.203 -0.356  0.033  0.365  2.113
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.84631    0.07205   11.75 < 2e-16 ***
## FirstAuthorFemale1 -0.00534    0.04578   -0.12  0.90721
## Year1997        0.16417    0.11333    1.45  0.14765
## Year1998        0.18027    0.09586    1.88  0.06021 .
## Year1999        0.26189    0.09866    2.65  0.00802 **
## Year2000        0.20007    0.09334    2.14  0.03223 *
## Year2001        0.35303    0.11677    3.02  0.00254 **
## Year2002        0.25541    0.08825    2.89  0.00385 **
## Year2003        0.08042    0.10322    0.78  0.43604
## Year2004        0.18386    0.09880    1.86  0.06293 .
## Year2005        0.17839    0.09239    1.93  0.05368 .
## Year2006        0.18299    0.08912    2.05  0.04021 *
```

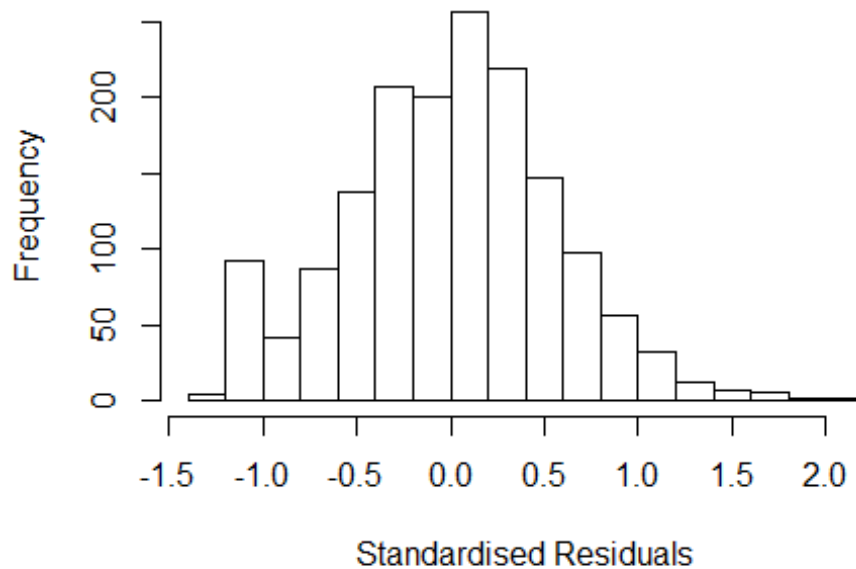


```

## Year2007          0.15619      0.08512      1.83  0.06672 .
## Year2008          0.19017      0.08804      2.16  0.03093 *
## Year2009          0.27422      0.08770      3.13  0.00180 **
## Year2010          0.35672      0.08919      4.00  6.6e-05 ***
## Year2011          0.31627      0.08814      3.59  0.00034 ***
## Year2012          0.28165      0.08690      3.24  0.00122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0246, Adjusted R-squared:  0.0142
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 131 weights are ~= 1. The remaining 1473 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.107  0.870  0.954  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.073 1      1.036
## Year              1.073 16      1.002

```

## Residuals from last author



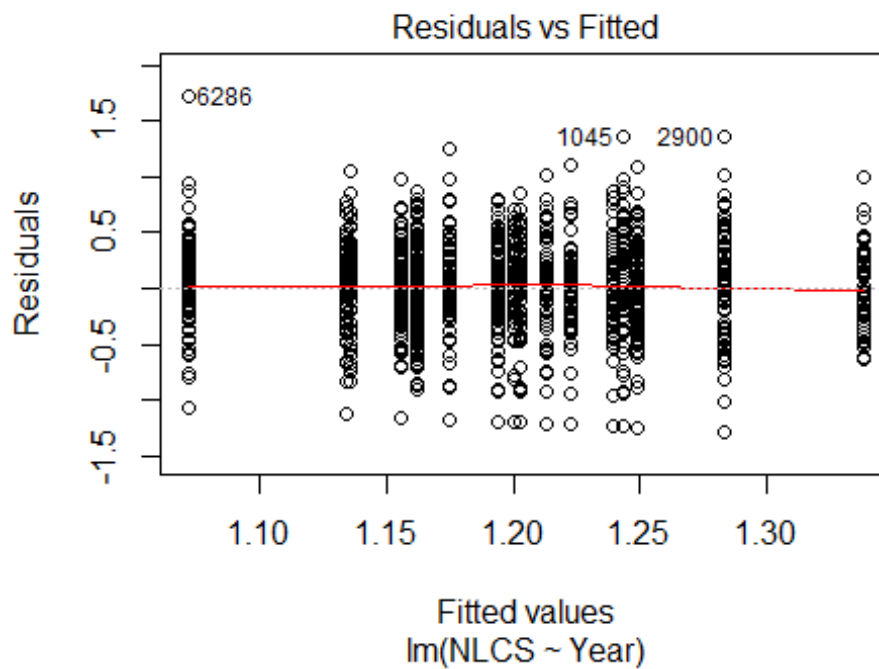
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2042 -0.3564 0.0343 0.3634 2.1122
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8481 0.0719 11.79 < 2e-16 ***
## LastAuthorFemale1 -0.0319 0.0589 -0.54 0.58803
## Year1997 0.1636 0.1132 1.45 0.14862
## Year1998 0.1799 0.0958 1.88 0.06054 .
## Year1999 0.2628 0.0982 2.68 0.00754 **
## Year2000 0.2017 0.0931 2.17 0.03042 *
## Year2001 0.3514 0.1174 2.99 0.00280 **
## Year2002 0.2561 0.0882 2.90 0.00373 **
## Year2003 0.0792 0.1033 0.77 0.44317
## Year2004 0.1827 0.0987 1.85 0.06444 .
## Year2005 0.1778 0.0920 1.93 0.05358 .
## Year2006 0.1828 0.0886 2.06 0.03931 *
```

```

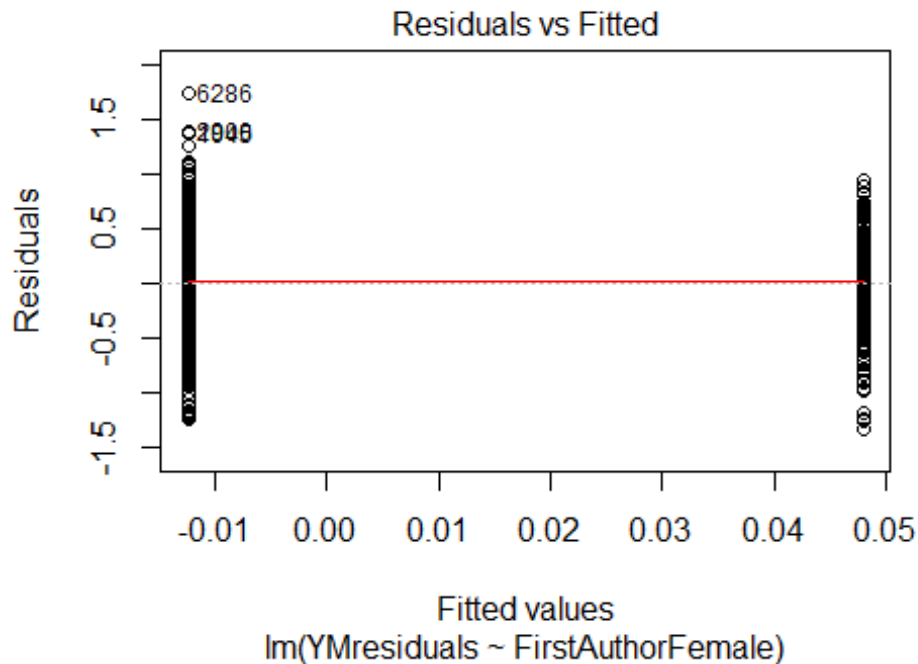
## Year2007          0.1565      0.0849      1.84  0.06548 .
## Year2008          0.1899      0.0877      2.17  0.03051 *
## Year2009          0.2756      0.0874      3.15  0.00165 **
## Year2010          0.3561      0.0890      4.00  6.6e-05 ***
## Year2011          0.3157      0.0875      3.61  0.00032 ***
## Year2012          0.2823      0.0869      3.25  0.00118 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0249, Adjusted R-squared:  0.0145
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 129 weights are ~= 1. The remaining 1475 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0996 0.8680 0.9530 0.9040 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.23e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1604"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3110"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 378 373 391 358 321 311 284 263 282 227 288 249 234 287 294
## 2011 2012
## 264 213
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 69 69 65 65 56 50 79 64 95 69 95 75 72 91 86
## 2011 2012

```

```
## 90 103
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 62 59 62 56 49 44 65 58 78 64 86 62 64 76 77
## 2011 2012
## 82 87
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 47, df = 16, p-value = 7e-05
```



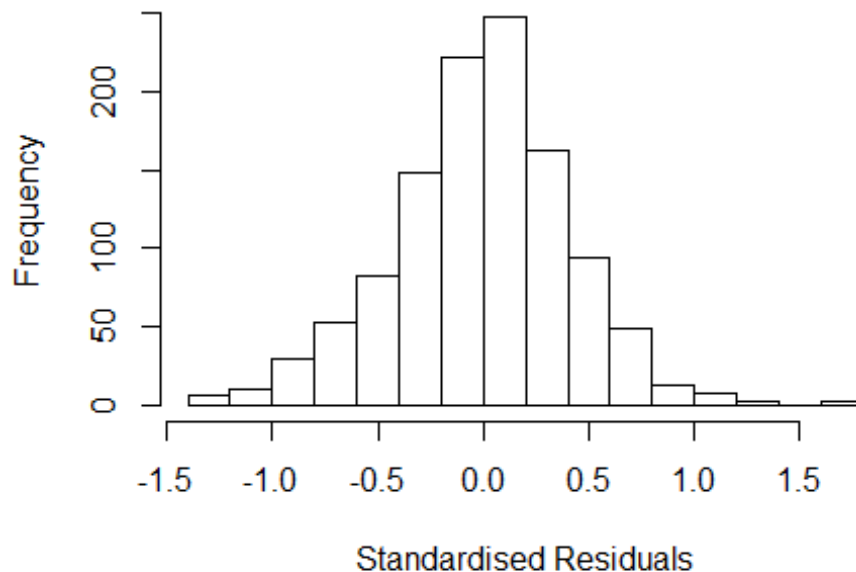
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 4.06492464801093"
## [1] "Male first author team size 2018 geometric mean: 3.22115348436374"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 510, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.38604688485739"
## [1] "Male last author team size 2018 geometric mean: 3.47017062375106"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.099	1	1.048
LastAuthorFemale	1.092	1	1.045
UniqueAuthors	1.471	4	1.049
Year	1.659	16	1.016

## Residuals from first and last author and team size



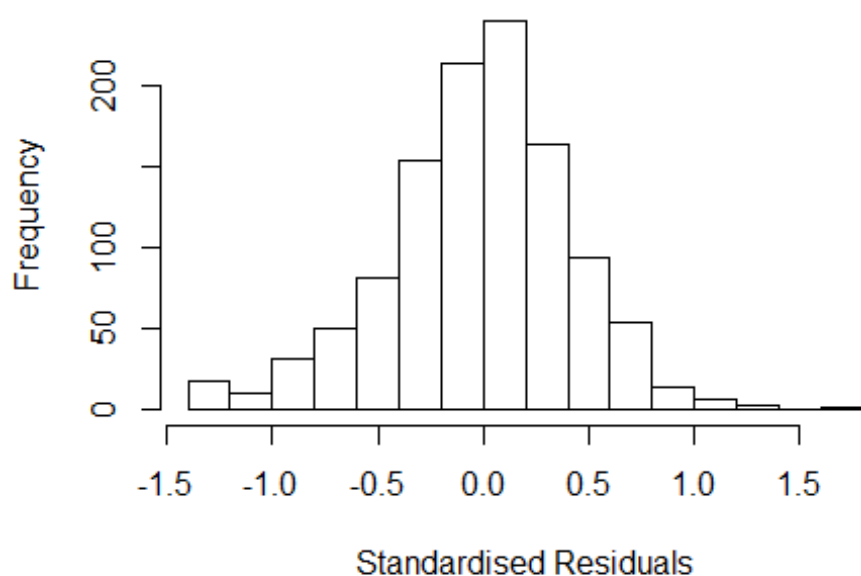
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.38651 -0.25534  0.00822  0.26147  1.79476
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94979    0.08001   11.87 < 2e-16 ***
## FirstAuthorFemale1 0.04291    0.03008    1.43  0.1540
## LastAuthorFemale1 -0.03668    0.04272   -0.86  0.3908
## UniqueAuthors2    0.30707    0.05740    5.35 1.1e-07 ***
## UniqueAuthors3    0.28483    0.05833    4.88 1.2e-06 ***
## UniqueAuthors4    0.39818    0.06192    6.43 1.9e-10 ***
## UniqueAuthors5    0.30963    0.05660    5.47 5.6e-08 ***
## Year1997          0.03854    0.07714    0.50  0.6175
## Year1998          0.05304    0.09361    0.57  0.5711
## Year1999         -0.04694    0.08885   -0.53  0.5974
```

```

## Year2000      -0.02257    0.09617   -0.23    0.8145
## Year2001      0.08359    0.08221    1.02    0.3095
## Year2002      0.05820    0.08802    0.66    0.5086
## Year2003     -0.00827    0.07610   -0.11    0.9135
## Year2004      0.02383    0.07660    0.31    0.7557
## Year2005     -0.10529    0.07256   -1.45    0.1470
## Year2006     -0.03101    0.07563   -0.41    0.6818
## Year2007      0.01216    0.07924    0.15    0.8781
## Year2008      0.03532    0.07487    0.47    0.6372
## Year2009     -0.05534    0.07304   -0.76    0.4488
## Year2010     -0.07131    0.07686   -0.93    0.3537
## Year2011     -0.07239    0.07606   -0.95    0.3414
## Year2012     -0.23638    0.07725   -3.06    0.0023 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.387
## Multiple R-squared:  0.1,   Adjusted R-squared:  0.0826
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 98 weights are ~= 1. The remaining 1033 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0004 0.8630 0.9490 0.8940 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.84e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.079 1 1.039
## LastAuthorFemale 1.073 1 1.036
## Year 1.151 16 1.004

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3201 -0.2644 0.0126 0.2600 1.7756
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.22021 0.06348 19.22 <2e-16 ***
## FirstAuthorFemale1 0.06956 0.03063 2.27 0.0233 *
## LastAuthorFemale1 -0.03545 0.04441 -0.80 0.4250
## Year1997 0.00720 0.07969 0.09 0.9280
## Year1998 0.08247 0.09780 0.84 0.3993
## Year1999 -0.07510 0.09318 -0.81 0.4204
## Year2000 -0.00997 0.10500 -0.09 0.9244
## Year2001 0.12058 0.08358 1.44 0.1494
## Year2002 0.06116 0.09348 0.65 0.5130
## Year2003 0.00528 0.07903 0.07 0.9468
## Year2004 0.05386 0.07835 0.69 0.4920
## Year2005 -0.08712 0.07486 -1.16 0.2448
```

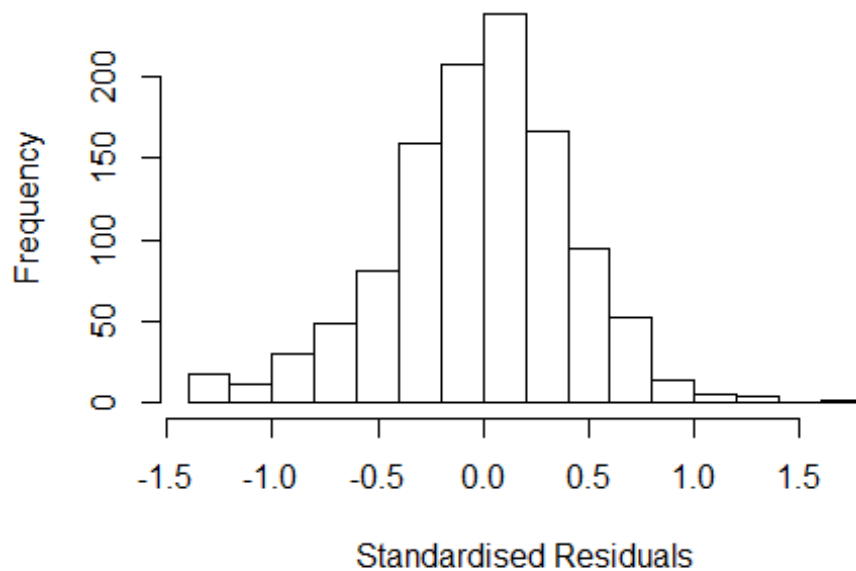


```

## Year2006      -0.00705    0.07824   -0.09    0.9282
## Year2007      0.02400    0.08221    0.29    0.7704
## Year2008      0.03036    0.07718    0.39    0.6941
## Year2009     -0.06568    0.07541   -0.87    0.3839
## Year2010     -0.06627    0.07777   -0.85    0.3943
## Year2011     -0.05835    0.07902   -0.74    0.4604
## Year2012     -0.20286    0.07771   -2.61    0.0092 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0354, Adjusted R-squared:  0.0198
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1034 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0054 0.8620 0.9500 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.078 1      1.038
## Year      1.078 16      1.002

```

## Residuals from first author



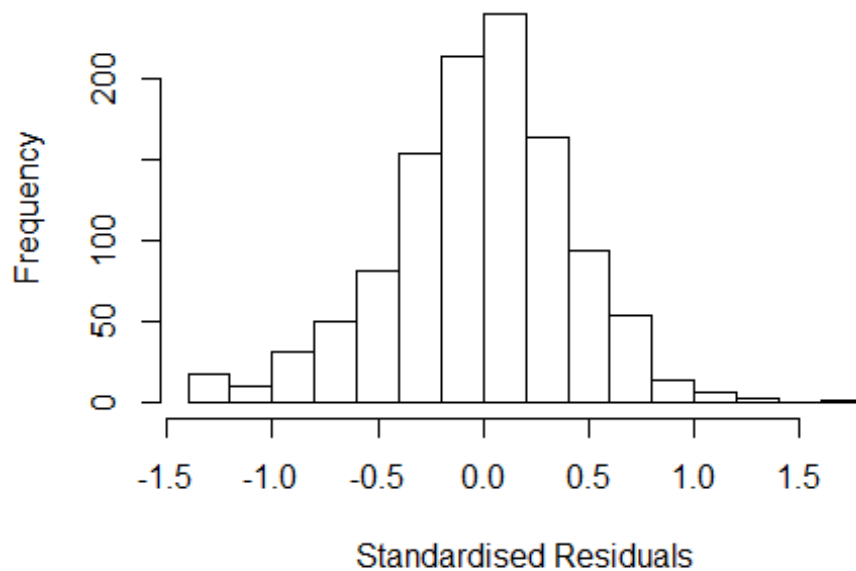
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3465 -0.2625 0.0152 0.2607 1.7788
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.21710 0.06333 19.22 <2e-16 ***
## FirstAuthorFemale1 0.06849 0.03079 2.22 0.0263 *
## Year1997 0.00771 0.07965 0.10 0.9229
## Year1998 0.08147 0.09802 0.83 0.4060
## Year1999 -0.07418 0.09316 -0.80 0.4261
## Year2000 -0.01138 0.10558 -0.11 0.9142
## Year2001 0.12292 0.08366 1.47 0.1421
## Year2002 0.06091 0.09357 0.65 0.5152
## Year2003 0.00463 0.07925 0.06 0.9535
## Year2004 0.05524 0.07820 0.71 0.4800
## Year2005 -0.08509 0.07469 -1.14 0.2548
## Year2006 -0.00779 0.07841 -0.10 0.9208
```

```

## Year2007          0.02240      0.08252      0.27      0.7861
## Year2008          0.02755      0.07725      0.36      0.7214
## Year2009         -0.06378      0.07544     -0.85      0.3980
## Year2010         -0.06538      0.07774     -0.84      0.4005
## Year2011         -0.05888      0.07898     -0.75      0.4561
## Year2012         -0.20293      0.07789     -2.61      0.0093 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.394
## Multiple R-squared:  0.0349, Adjusted R-squared:  0.0201
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 97 weights are ~= 1. The remaining 1034 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0052 0.8610 0.9500 0.8910 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.074 1      1.036
## Year      1.074 16      1.002

```

## Residuals from last author



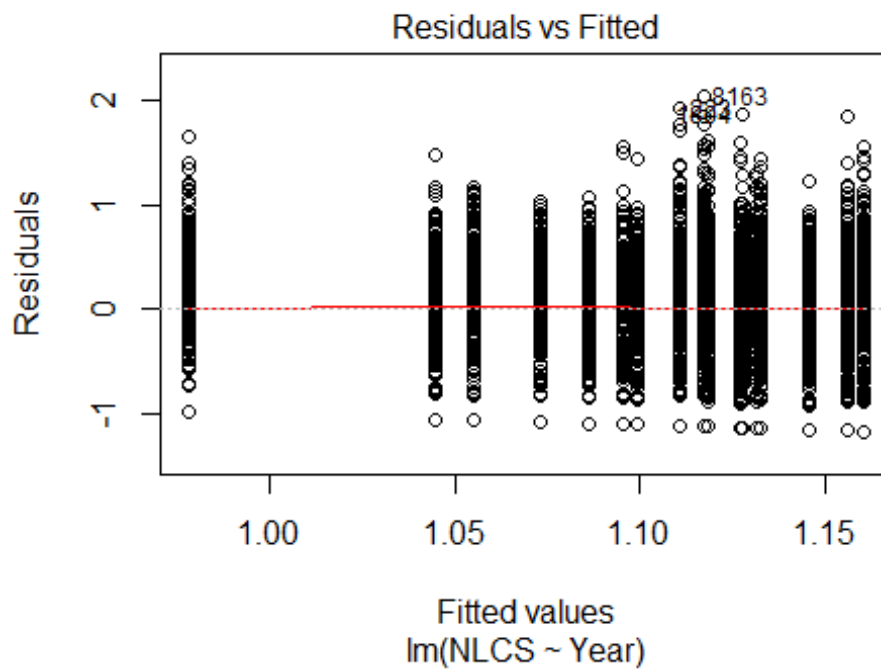
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3134 -0.2640 0.0119 0.2650 1.7534
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2291 0.0631 19.47 <2e-16 ***
## LastAuthorFemale1 -0.0312 0.0443 -0.70 0.482
## Year1997 0.0068 0.0798 0.09 0.932
## Year1998 0.0843 0.0977 0.86 0.388
## Year1999 -0.0772 0.0935 -0.83 0.409
## Year2000 -0.0087 0.1049 -0.08 0.934
## Year2001 0.1219 0.0836 1.46 0.145
## Year2002 0.0553 0.0931 0.59 0.553
## Year2003 0.0112 0.0786 0.14 0.886
## Year2004 0.0616 0.0777 0.79 0.428
## Year2005 -0.0792 0.0751 -1.05 0.292
## Year2006 0.0030 0.0776 0.04 0.969
```

```

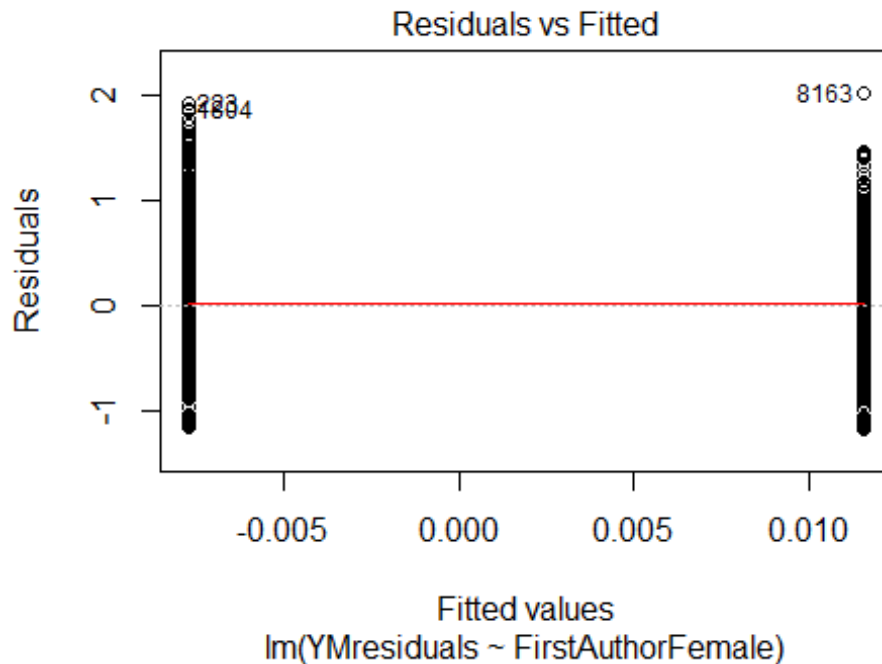
## Year2007          0.0262      0.0818      0.32      0.748
## Year2008          0.0284      0.0774      0.37      0.714
## Year2009         -0.0509      0.0746     -0.68      0.495
## Year2010         -0.0584      0.0777     -0.75      0.453
## Year2011         -0.0512      0.0789     -0.65      0.516
## Year2012         -0.1895      0.0778     -2.44      0.015 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.395
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.016
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 100 weights are ~= 1. The remaining 1031 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0104 0.8580 0.9490 0.8910 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.84e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1131"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3200"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 375 324 305 309 332 388 331 308 377 377 461 527 520 449 430
## 2011 2012
## 491 467
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 300 248 234 220 226 248 252 254 310 315 371 447 431 399 364
## 2011 2012

```

```
## 426 408
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 282 230 221 205 209 223 233 221 278 289 327 402 397 364 338
## 2011 2012
## 375 382
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 97, df = 16, p-value = 1e-13
```

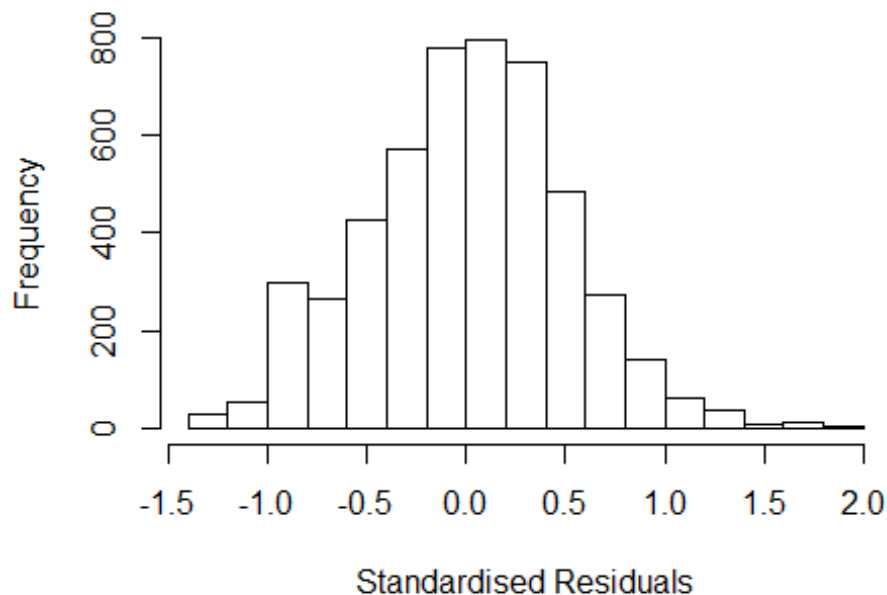


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 20, df = 1, p-value = 9e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 2.46939468980484"
## [1] "Male first author team size 2018 geometric mean: 2.33395664285975"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.44079178453066"
## [1] "Male last author team size 2018 geometric mean: 2.37374156708038"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.077 1          1.038
## LastAuthorFemale  1.054 1          1.027
## UniqueAuthors    1.151 4          1.018
## Year              1.170 16         1.005
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3969 -0.3255  0.0154  0.3363  1.9343
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.93063    0.03916   23.76  <2e-16 ***
## FirstAuthorFemale1 0.01558    0.01492    1.04   0.296
## LastAuthorFemale1 -0.01654    0.01516   -1.09   0.275
## UniqueAuthors2    0.27079    0.01970   13.75  <2e-16 ***
## UniqueAuthors3    0.33294    0.02167   15.37  <2e-16 ***
## UniqueAuthors4    0.38710    0.02625   14.75  <2e-16 ***
## UniqueAuthors5    0.42009    0.02958   14.20  <2e-16 ***
## Year1997          0.03062    0.05008    0.61   0.541
## Year1998         -0.00584    0.05023   -0.12   0.907
## Year1999         -0.05371    0.05082   -1.06   0.291
```

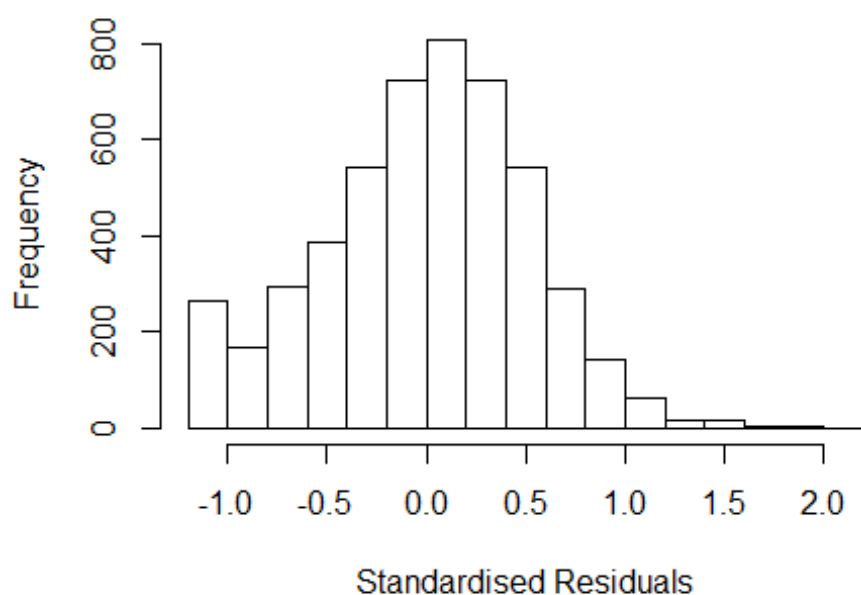


```

## Year2000      0.03048    0.05258    0.58    0.562
## Year2001     -0.11407    0.05572   -2.05    0.041 *
## Year2002     -0.05781    0.05272   -1.10    0.273
## Year2003     -0.07191    0.05133   -1.40    0.161
## Year2004     -0.02442    0.04699   -0.52    0.603
## Year2005     -0.08767    0.04560   -1.92    0.055 .
## Year2006     -0.04343    0.04422   -0.98    0.326
## Year2007     -0.03697    0.04250   -0.87    0.384
## Year2008     -0.05156    0.04395   -1.17    0.241
## Year2009      0.00668    0.04513    0.15    0.882
## Year2010     -0.00316    0.04530   -0.07    0.944
## Year2011     -0.00385    0.04612   -0.08    0.934
## Year2012     -0.06641    0.04754   -1.40    0.162
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.489
## Multiple R-squared:  0.0915, Adjusted R-squared:  0.0875
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 409 weights are ~= 1. The remaining 4567 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0828 0.8640 0.9500 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.01e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1 1.029
## LastAuthorFemale 1.046 1 1.023
## Year 1.029 16 1.001

```

## Residuals from first and last author



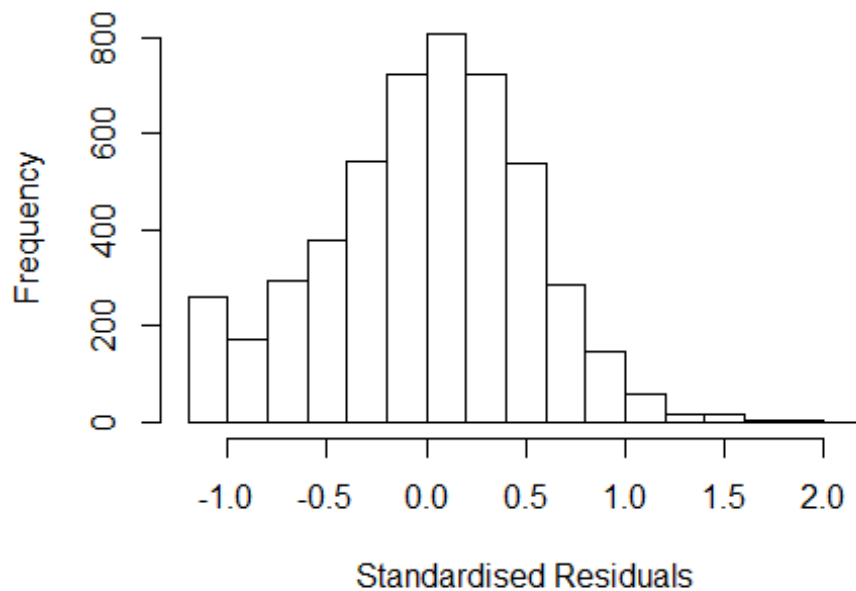
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1909 -0.3402 0.0259 0.3513 2.0087
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09955 0.03916 28.08 <2e-16 ***
## FirstAuthorFemale1 0.03732 0.01552 2.40 0.016 *
## LastAuthorFemale1 -0.01772 0.01585 -1.12 0.264
## Year1997 0.02227 0.05257 0.42 0.672
## Year1998 -0.00692 0.05206 -0.13 0.894
## Year1999 -0.07432 0.05261 -1.41 0.158
## Year2000 0.03504 0.05513 0.64 0.525
## Year2001 -0.12937 0.05761 -2.25 0.025 *
## Year2002 -0.03647 0.05636 -0.65 0.518
## Year2003 -0.03166 0.05483 -0.58 0.564
## Year2004 0.00945 0.05053 0.19 0.852
## Year2005 -0.00970 0.04793 -0.20 0.840
```

```

## Year2006          0.01968      0.04596      0.43      0.668
## Year2007          0.03655      0.04395      0.83      0.406
## Year2008          0.01762      0.04581      0.38      0.701
## Year2009          0.05401      0.04758      1.14      0.256
## Year2010          0.03346      0.04767      0.70      0.483
## Year2011          0.06714      0.04803      1.40      0.162
## Year2012          0.00142      0.04954      0.03      0.977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.00818,    Adjusted R-squared:  0.00458
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 416 weights are ~= 1. The remaining 4560 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0817 0.8640 0.9490 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.011
## Year              1.023 16      1.001

```

## Residuals from first author



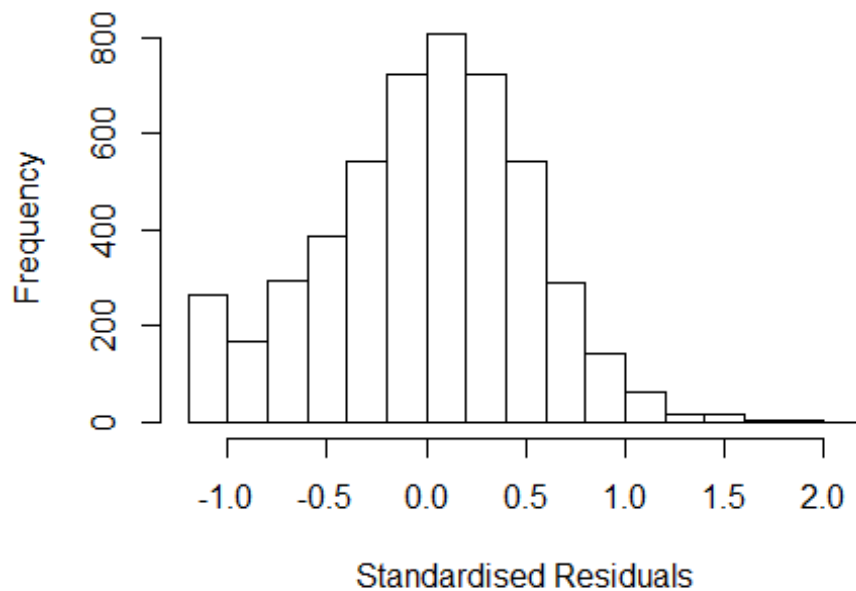
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1930 -0.3393 0.0245 0.3511 2.0193
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.095471 0.038997 28.09 <2e-16 ***
## FirstAuthorFemale1 0.031567 0.015424 2.05 0.041 *
## Year1997 0.022054 0.052681 0.42 0.676
## Year1998 -0.007316 0.052138 -0.14 0.888
## Year1999 -0.073086 0.052700 -1.39 0.166
## Year2000 0.034976 0.055181 0.63 0.526
## Year2001 -0.129661 0.057575 -2.25 0.024 *
## Year2002 -0.036045 0.056396 -0.64 0.523
## Year2003 -0.031317 0.054859 -0.57 0.568
## Year2004 0.010018 0.050575 0.20 0.843
## Year2005 -0.010238 0.048001 -0.21 0.831
## Year2006 0.019737 0.046056 0.43 0.668
```

```

## Year2007          0.036486    0.044060    0.83    0.408
## Year2008          0.016971    0.045882    0.37    0.711
## Year2009          0.053624    0.047679    1.12    0.261
## Year2010          0.033298    0.047730    0.70    0.485
## Year2011          0.065972    0.048134    1.37    0.171
## Year2012          0.000661    0.049672    0.01    0.989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.00798,    Adjusted R-squared:  0.00458
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 408 weights are ~= 1. The remaining 4568 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0772 0.8640 0.9500 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.01 1          1.005
## Year            1.01 16          1.000

```

## Residuals from last author



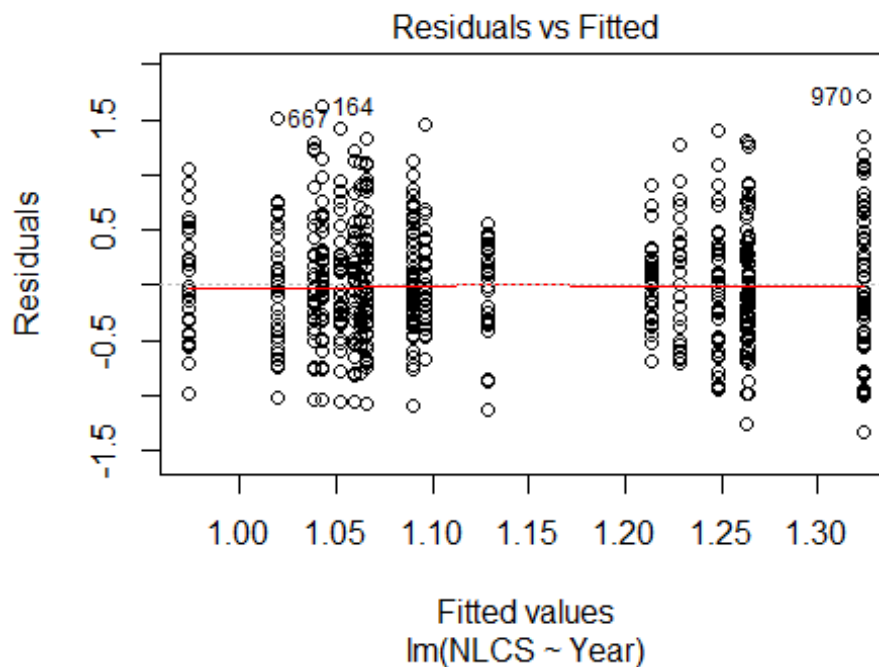
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1778 -0.3424  0.0293  0.3505  2.0341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.10922    0.03883   28.56  <2e-16 ***
## LastAuthorFemale1 -0.00523    0.01570   -0.33    0.739
## Year1997         0.02134    0.05261    0.41    0.685
## Year1998        -0.00832    0.05204   -0.16    0.873
## Year1999        -0.07241    0.05267   -1.37    0.169
## Year2000         0.03556    0.05521    0.64    0.520
## Year2001        -0.13246    0.05759   -2.30    0.021 *
## Year2002        -0.03545    0.05653   -0.63    0.531
## Year2003        -0.03226    0.05495   -0.59    0.557
## Year2004         0.00927    0.05059    0.18    0.855
## Year2005        -0.00833    0.04805   -0.17    0.862
## Year2006         0.02371    0.04598    0.52    0.606
```

```

## Year2007      0.03931    0.04399    0.89    0.372
## Year2008      0.01844    0.04587    0.40    0.688
## Year2009      0.05565    0.04767    1.17    0.243
## Year2010      0.03458    0.04771    0.72    0.469
## Year2011      0.06863    0.04811    1.43    0.154
## Year2012      0.00368    0.04962    0.07    0.941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.00712,    Adjusted R-squared:  0.00372
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 414 weights are ~= 1. The remaining 4562 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0717 0.8650 0.9490 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4976"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3201"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   39   34   49   64   56   44   45   43   32   30   41   65   52   77   55
## 2011 2012
##   66   75
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   28   35   48   35   33   34   34   28   27   36   49   38   64   42
## 2011 2012

```

```
## 58 66
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33 25 34 47 34 32 34 33 24 24 34 46 33 59 41
## 2011 2012
## 55 62
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 30, df = 16, p-value = 0.02
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.014, df = 1, p-value = 0.9

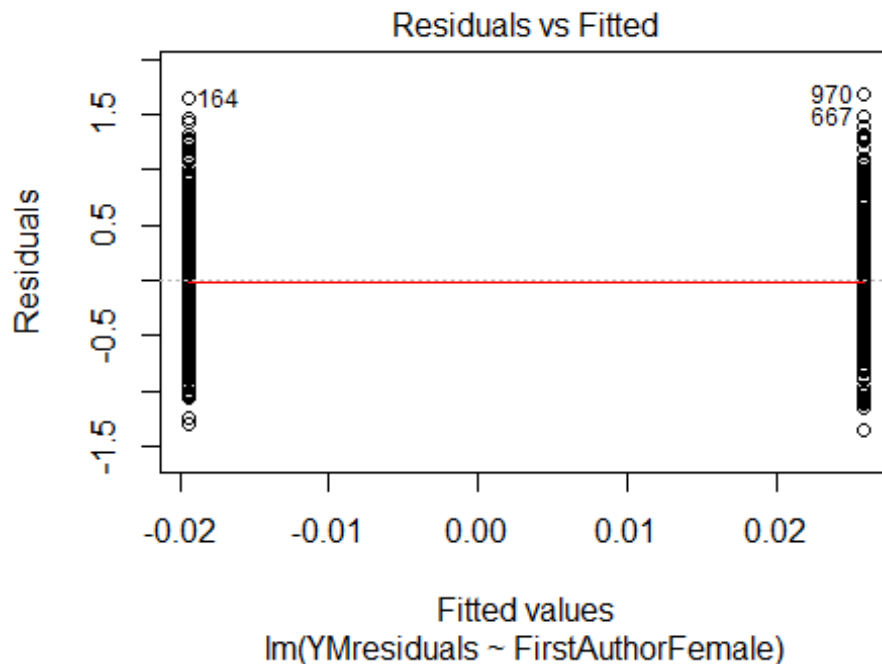
## [1] "Female first author team size 2018 geometric mean: 2.84731468735758"
## [1] "Male first author team size 2018 geometric mean: 1.60831190760079"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



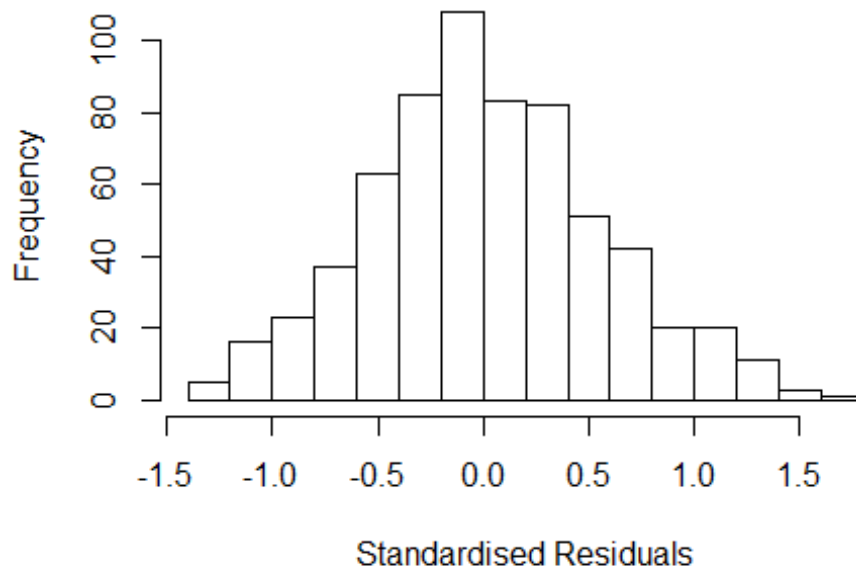
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 350, p-value = 7e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.55473659852179"
## [1] "Male last author team size 2018 geometric mean: 1.94490100403462"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 280, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.362 1          1.167
## LastAuthorFemale 1.360 1          1.166
## UniqueAuthors    1.800 4          1.076
## Year              1.856 16         1.020
```

## Residuals from first and last author and team size



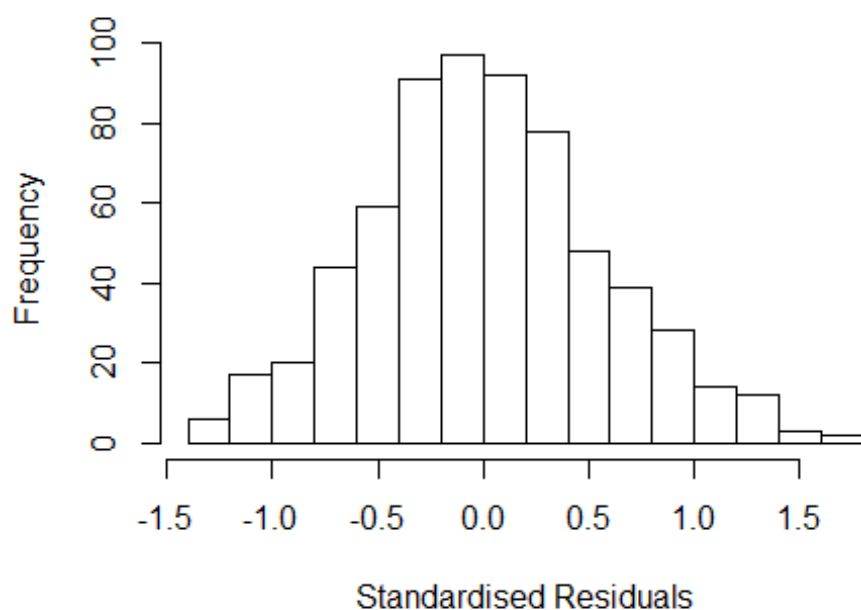
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3543 -0.3533 -0.0321 0.3441 1.6340
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94972 0.10140 9.37 <2e-16 ***
## FirstAuthorFemale1 -0.00313 0.05034 -0.06 0.950
## LastAuthorFemale1 0.04465 0.05082 0.88 0.380
## UniqueAuthors2 0.14699 0.06010 2.45 0.015 *
## UniqueAuthors3 0.10349 0.06966 1.49 0.138
## UniqueAuthors4 -0.09038 0.07814 -1.16 0.248
## UniqueAuthors5 0.07885 0.09032 0.87 0.383
## Year1997 0.16219 0.12089 1.34 0.180
## Year1998 0.01929 0.13808 0.14 0.889
## Year1999 0.03150 0.13791 0.23 0.819
```

```

## Year2000          0.04742      0.13421      0.35      0.724
## Year2001         -0.06481      0.13073     -0.50      0.620
## Year2002          0.23061      0.13739      1.68      0.094 .
## Year2003          0.10080      0.12314      0.82      0.413
## Year2004          0.01376      0.14314      0.10      0.923
## Year2005          0.19775      0.16362      1.21      0.227
## Year2006          0.08874      0.11855      0.75      0.454
## Year2007          0.21108      0.12389      1.70      0.089 .
## Year2008         -0.04797      0.14179     -0.34      0.735
## Year2009          0.08215      0.12001      0.68      0.494
## Year2010          0.19091      0.13355      1.43      0.153
## Year2011          0.03079      0.12826      0.24      0.810
## Year2012          0.26076      0.15024      1.74      0.083 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.0519, Adjusted R-squared:  0.0187
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 48 weights are ~= 1. The remaining 602 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.335  0.866  0.955  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.54e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.279 1      1.131
## LastAuthorFemale  1.254 1      1.120
## Year              1.202 16      1.006

```

## Residuals from first and last author



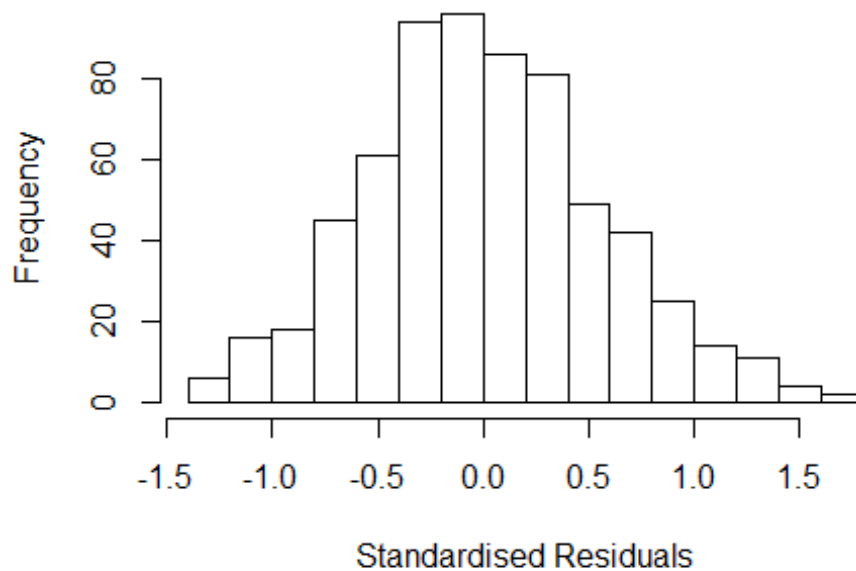
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3368 -0.3506 -0.0194 0.3334 1.6962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9813 0.0993 9.88 <2e-16 ***
## FirstAuthorFemale1 0.0171 0.0497 0.34 0.731
## LastAuthorFemale1 0.0535 0.0505 1.06 0.290
## Year1997 0.1734 0.1199 1.45 0.148
## Year1998 0.0335 0.1424 0.23 0.814
## Year1999 0.0284 0.1376 0.21 0.836
## Year2000 0.0670 0.1330 0.50 0.614
## Year2001 -0.0649 0.1313 -0.49 0.621
## Year2002 0.2557 0.1340 1.91 0.057 .
## Year2003 0.1304 0.1190 1.10 0.274
## Year2004 0.0226 0.1422 0.16 0.874
## Year2005 0.2169 0.1620 1.34 0.181
```

```

## Year2006          0.0828      0.1195      0.69      0.489
## Year2007          0.2322      0.1254      1.85      0.065 .
## Year2008          0.0107      0.1354      0.08      0.937
## Year2009          0.0873      0.1180      0.74      0.460
## Year2010          0.1999      0.1340      1.49      0.136
## Year2011          0.0585      0.1271      0.46      0.646
## Year2012          0.2849      0.1483      1.92      0.055 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0354, Adjusted R-squared:  0.0079
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 53 weights are ~= 1. The remaining 597 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.310  0.868  0.955  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.54e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.085 1      1.042
## Year              1.085 16      1.003

```

## Residuals from first author

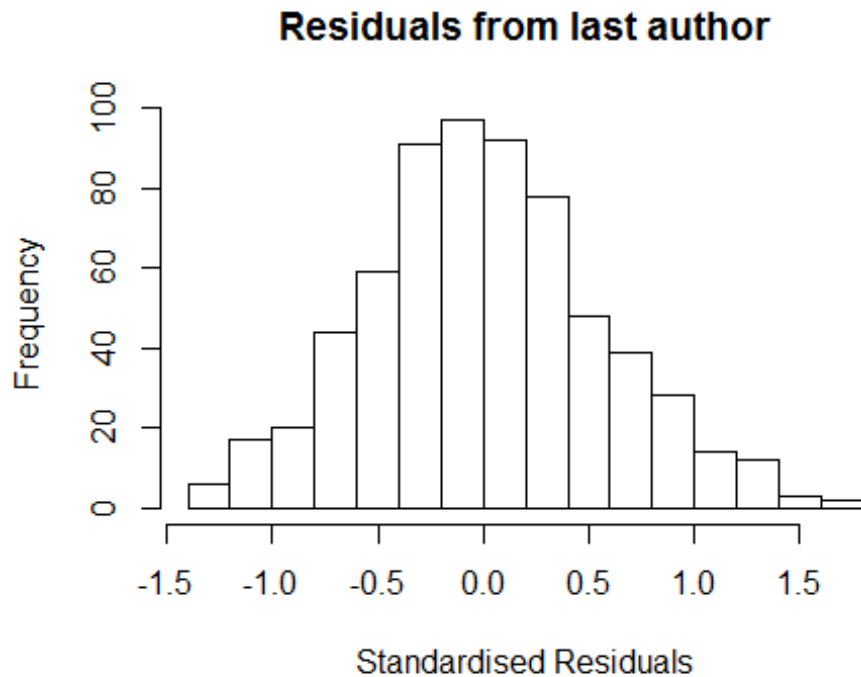


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3142 -0.3571 -0.0171 0.3432 1.7188
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.99229 0.09787 10.14 <2e-16 ***
## FirstAuthorFemale1 0.03972 0.04599 0.86 0.388
## Year1997 0.17277 0.12023 1.44 0.151
## Year1998 0.02903 0.14203 0.20 0.838
## Year1999 0.02661 0.13704 0.19 0.846
## Year2000 0.07124 0.13332 0.53 0.593
## Year2001 -0.06452 0.13250 -0.49 0.626
## Year2002 0.25675 0.13549 1.89 0.059 .
## Year2003 0.12740 0.11834 1.08 0.282
## Year2004 0.02353 0.14267 0.16 0.869
## Year2005 0.21419 0.16144 1.33 0.185
## Year2006 0.08604 0.12015 0.72 0.474
```

```

## Year2007          0.23551    0.12544    1.88    0.061 .
## Year2008          0.00499    0.13540    0.04    0.971
## Year2009          0.09109    0.11856    0.77    0.443
## Year2010          0.19730    0.13381    1.47    0.141
## Year2011          0.05608    0.12700    0.44    0.659
## Year2012          0.28223    0.14797    1.91    0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0335, Adjusted R-squared:  0.00745
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 52 weights are ~= 1. The remaining 598 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.299  0.866  0.953  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.54e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.068 1      1.033
## Year              1.068 16      1.002

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3317 -0.3444 -0.0174  0.3365  1.7013
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9843     0.0990   9.94 <2e-16 ***
## LastAuthorFemale1  0.0609     0.0468   1.30  0.193
## Year1997          0.1712     0.1198   1.43  0.153
## Year1998          0.0354     0.1418   0.25  0.803
## Year1999          0.0296     0.1377   0.22  0.830
## Year2000          0.0691     0.1333   0.52  0.604
## Year2001         -0.0641     0.1314  -0.49  0.626
## Year2002          0.2575     0.1336   1.93  0.054 .
## Year2003          0.1333     0.1191   1.12  0.263
## Year2004          0.0247     0.1422   0.17  0.862
## Year2005          0.2187     0.1614   1.35  0.176
## Year2006          0.0830     0.1196   0.69  0.488
```

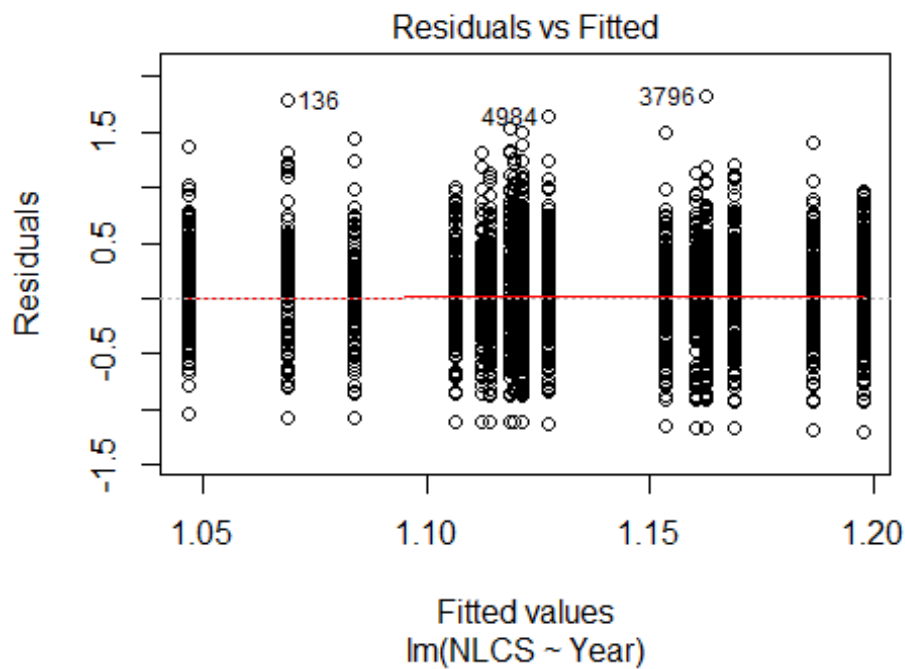


```

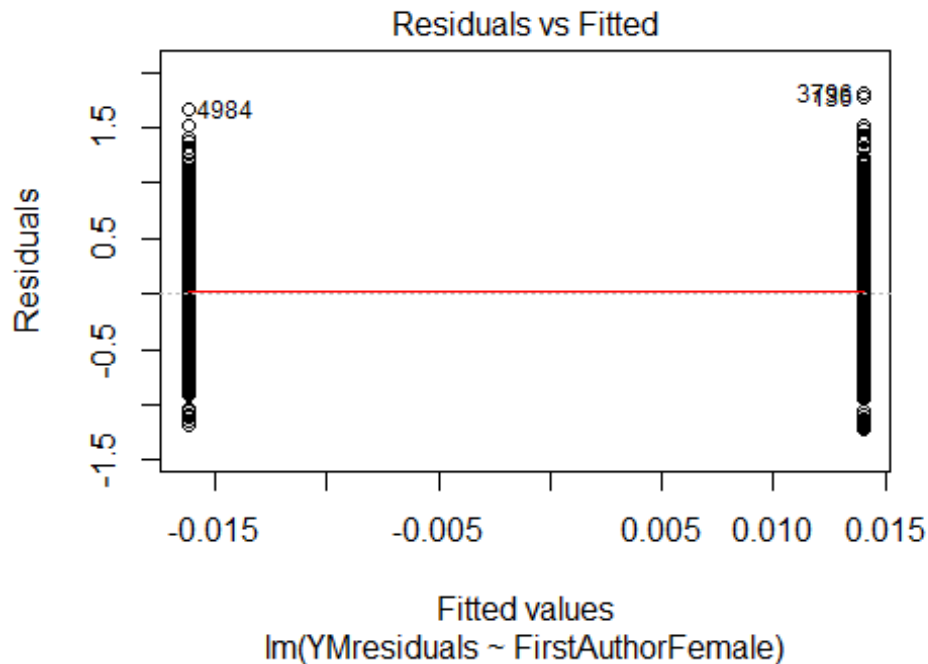
## Year2007          0.2336      0.1256      1.86      0.063 .
## Year2008          0.0134      0.1352      0.10      0.921
## Year2009          0.0880      0.1179      0.75      0.456
## Year2010          0.2046      0.1328      1.54      0.124
## Year2011          0.0612      0.1271      0.48      0.631
## Year2012          0.2865      0.1484      1.93      0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.0353, Adjusted R-squared:  0.00931
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 50 weights are ~= 1. The remaining 600 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.307  0.869  0.954  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.54e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 650"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3202"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 256 240 221 195 271 264 248 238 225 209 256 320 318 387 416
## 2011 2012
## 420 469
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 189 176 143 138 163 135 163 183 191 187 214 284 262 295 329
## 2011 2012

```

```
## 328 363
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 176 168 140 129 148 126 150 151 162 171 197 262 244 275 308
## 2011 2012
## 300 340
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 81, df = 16, p-value = 1e-10
```

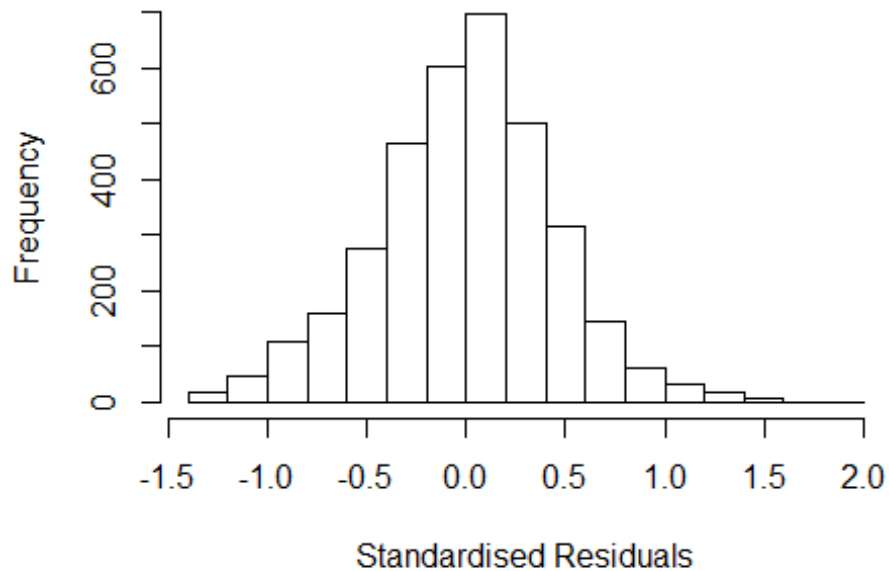


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.3, df = 1, p-value = 0.07
```



```
## [1] "Female first author team size 2018 geometric mean: 2.70231060440917"
## [1] "Male first author team size 2018 geometric mean: 2.28844696145813"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 24000, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.58920642099493"
## [1] "Male last author team size 2018 geometric mean: 2.44424513535331"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 22000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.128 1      1.062
## LastAuthorFemale  1.123 1      1.060
## UniqueAuthors     1.241 4      1.027
## Year              1.312 16     1.009
```

## Residuals from first and last author and team size



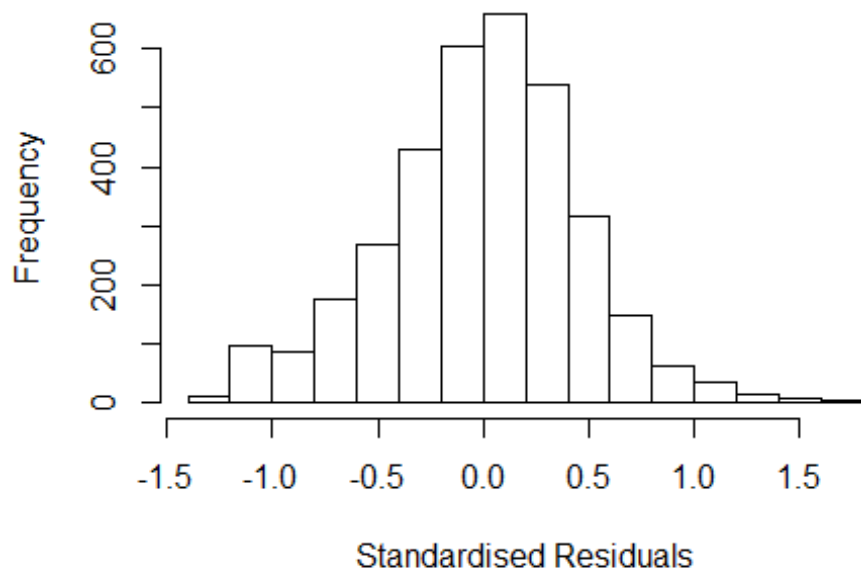
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3802 -0.2734  0.0131  0.2840  1.8942
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.963786   0.040159   24.00  < 2e-16 ***
## FirstAuthorFemale1 -0.034167   0.015901   -2.15   0.032 *
## LastAuthorFemale1 -0.041131   0.016244   -2.53   0.011 *
## UniqueAuthors2     0.214801   0.022350    9.61  < 2e-16 ***
## UniqueAuthors3     0.242743   0.024065   10.09  < 2e-16 ***
## UniqueAuthors4     0.232040   0.028233    8.22  2.9e-16 ***
## UniqueAuthors5     0.328339   0.028649   11.46  < 2e-16 ***
## Year1997          -0.033809   0.051335   -0.66   0.510
## Year1998           0.043140   0.060986    0.71   0.479
## Year1999           0.020633   0.056107    0.37   0.713
```

```

## Year2000      -0.024953    0.053621   -0.47    0.642
## Year2001      0.057950    0.058519    0.99    0.322
## Year2002      0.069414    0.057176    1.21    0.225
## Year2003      0.067123    0.057117    1.18    0.240
## Year2004      0.062349    0.050306    1.24    0.215
## Year2005      0.002200    0.048994    0.04    0.964
## Year2006     -0.007356    0.048594   -0.15    0.880
## Year2007      0.029970    0.044633    0.67    0.502
## Year2008      0.040562    0.048087    0.84    0.399
## Year2009      0.088108    0.046692    1.89    0.059 .
## Year2010      0.000671    0.047075    0.01    0.989
## Year2011      0.020699    0.046571    0.44    0.657
## Year2012      0.019763    0.049813    0.40    0.692
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0724, Adjusted R-squared:  0.0664
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 309 weights are ~= 1. The remaining 3138 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0019 0.8620 0.9480 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.90e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.098 1          1.048
## LastAuthorFemale 1.096 1          1.047
## Year      1.080 16          1.002

```

## Residuals from first and last author



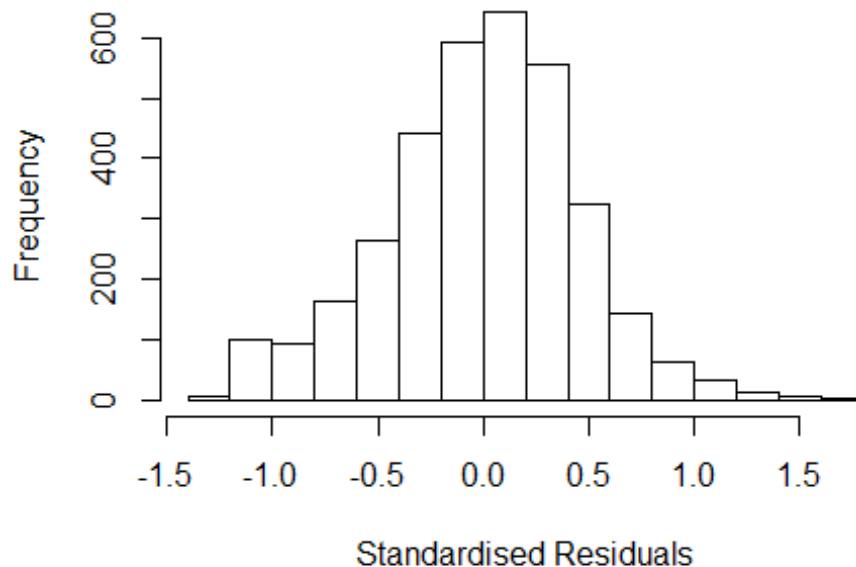
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2458 -0.2908  0.0177  0.2921  1.7838
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.08285    0.04202   25.77 < 2e-16 ***
## FirstAuthorFemale1 -0.01986    0.01620   -1.23  0.22040
## LastAuthorFemale1 -0.05135    0.01659   -3.09  0.00199 **
## Year1997        -0.01153    0.05404   -0.21  0.83111
## Year1998         0.05725    0.06299    0.91  0.36344
## Year1999         0.04065    0.05964    0.68  0.49559
## Year2000        -0.00333    0.05475   -0.06  0.95153
## Year2001         0.08319    0.06166    1.35  0.17741
## Year2002         0.12134    0.06015    2.02  0.04372 *
## Year2003         0.13901    0.05915    2.35  0.01882 *
## Year2004         0.12221    0.05298    2.31  0.02112 *
## Year2005         0.07506    0.05148    1.46  0.14492
```

```

## Year2006          0.05662    0.05086    1.11  0.26575
## Year2007          0.10037    0.04688    2.14  0.03235 *
## Year2008          0.11638    0.05017    2.32  0.02042 *
## Year2009          0.16299    0.04888    3.33  0.00086 ***
## Year2010          0.07277    0.04929    1.48  0.13989
## Year2011          0.08965    0.04924    1.82  0.06874 .
## Year2012          0.10732    0.05118    2.10  0.03608 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared:  0.0149, Adjusted R-squared:  0.00969
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 280 weights are ~= 1. The remaining 3167 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.040  0.863   0.949   0.894   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.90e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year              1.042 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2317 -0.2894 0.0173 0.2870 1.7925
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.06966 0.04176 25.62 < 2e-16 ***
## FirstAuthorFemale1 -0.03567 0.01587 -2.25 0.02461 *
## Year1997 -0.00948 0.05392 -0.18 0.86043
## Year1998 0.05237 0.06257 0.84 0.40270
## Year1999 0.04159 0.05955 0.70 0.48498
## Year2000 -0.00223 0.05479 -0.04 0.96752
## Year2001 0.08505 0.06149 1.38 0.16668
## Year2002 0.12285 0.06046 2.03 0.04223 *
## Year2003 0.14425 0.05913 2.44 0.01475 *
## Year2004 0.12407 0.05297 2.34 0.01924 *
## Year2005 0.07414 0.05139 1.44 0.14923
## Year2006 0.05681 0.05092 1.12 0.26469
```

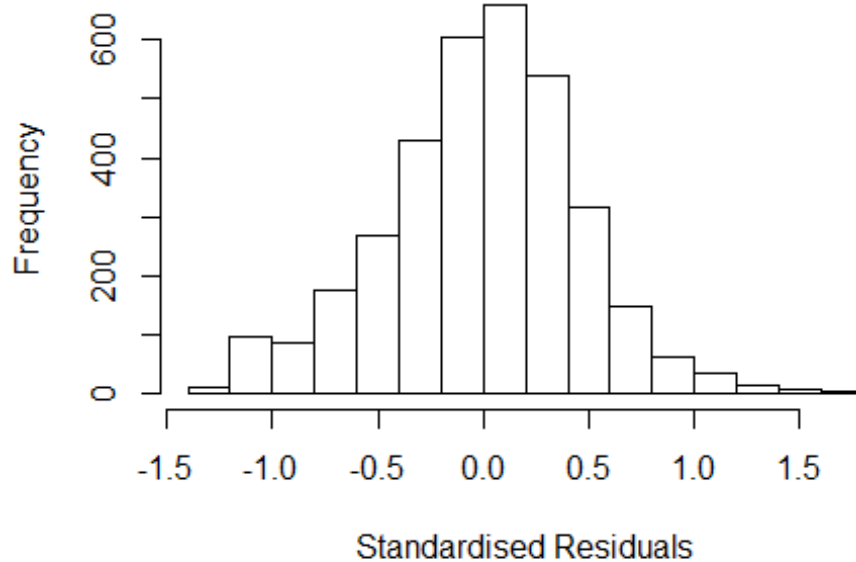


```

## Year2007          0.10152      0.04678      2.17  0.03007 *
## Year2008          0.12085      0.05011      2.41  0.01593 *
## Year2009          0.16200      0.04883      3.32  0.00092 ***
## Year2010          0.07173      0.04923      1.46  0.14517
## Year2011          0.08854      0.04932      1.80  0.07271 .
## Year2012          0.10635      0.05110      2.08  0.03748 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared:  0.012, Adjusted R-squared:  0.00712
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 288 weights are ~= 1. The remaining 3159 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0368 0.8630 0.9500 0.8930 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.90e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.041 1          1.020
## Year            1.041 16          1.001

```

## Residuals from last author



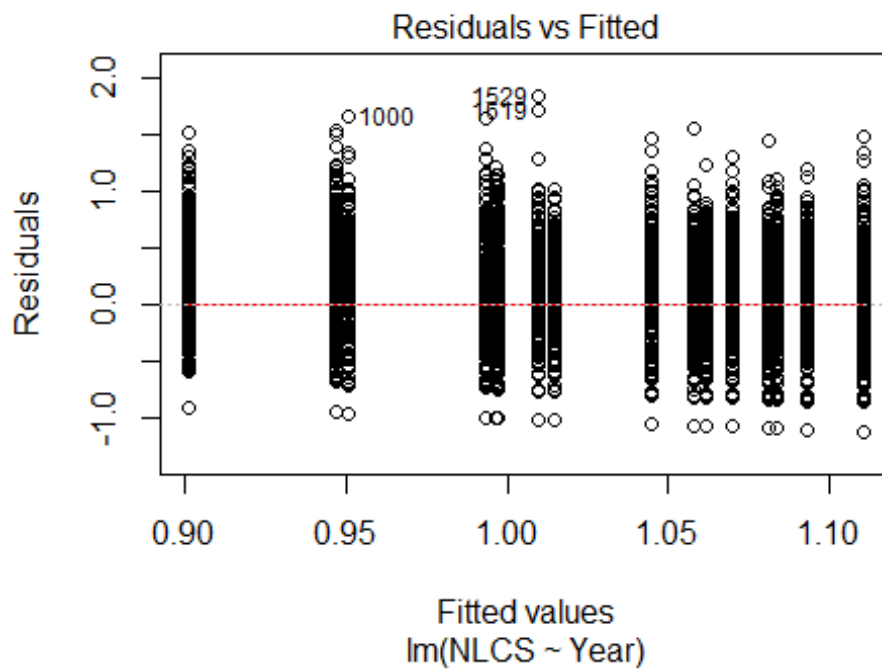
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2384 -0.2914  0.0164  0.2908  1.7927
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.07823    0.04180   25.80 < 2e-16 ***
## LastAuthorFemale1 -0.05771    0.01624   -3.55 0.00039 ***
## Year1997       -0.01106    0.05398   -0.20 0.83772
## Year1998        0.05586    0.06296    0.89 0.37502
## Year1999        0.03948    0.05969    0.66 0.50840
## Year2000       -0.00468    0.05469   -0.09 0.93182
## Year2001        0.08126    0.06147    1.32 0.18632
## Year2002        0.11908    0.05993    1.99 0.04699 *
## Year2003        0.13618    0.05906    2.31 0.02119 *
## Year2004        0.11906    0.05288    2.25 0.02440 *
## Year2005        0.07352    0.05146    1.43 0.15316
## Year2006        0.05456    0.05079    1.07 0.28278
```

```

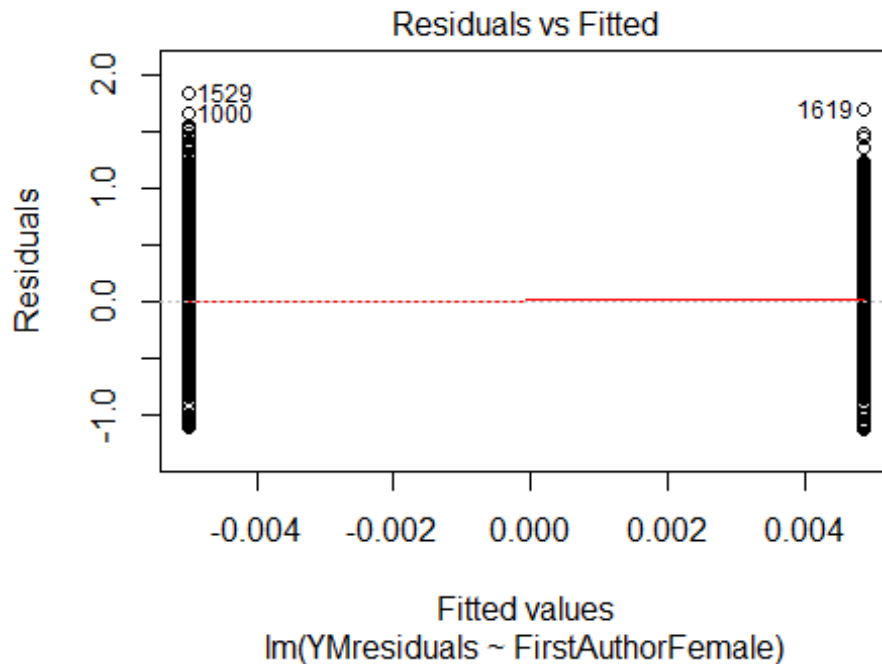
## Year2007      0.09839      0.04683      2.10  0.03570 *
## Year2008      0.11209      0.04991      2.25  0.02478 *
## Year2009      0.16020      0.04876      3.29  0.00103 **
## Year2010      0.07051      0.04927      1.43  0.15252
## Year2011      0.08643      0.04912      1.76  0.07856 .
## Year2012      0.10399      0.05109      2.04  0.04188 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.426
## Multiple R-squared:  0.0143, Adjusted R-squared:  0.00944
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 284 weights are ~= 1. The remaining 3163 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0376 0.8630 0.9490 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.90e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3447"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3203"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 318 278 300 286 321 352 351 311 350 356 383 484 421 553 595
## 2011 2012
## 650 657
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 232 206 226 229 217 232 280 245 301 300 313 424 370 472 506
## 2011 2012

```

```
## 567 551
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 215 196 210 217 199 217 253 213 267 273 292 374 330 435 472
## 2011 2012
## 534 511
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 75, df = 16, p-value = 1e-09
```

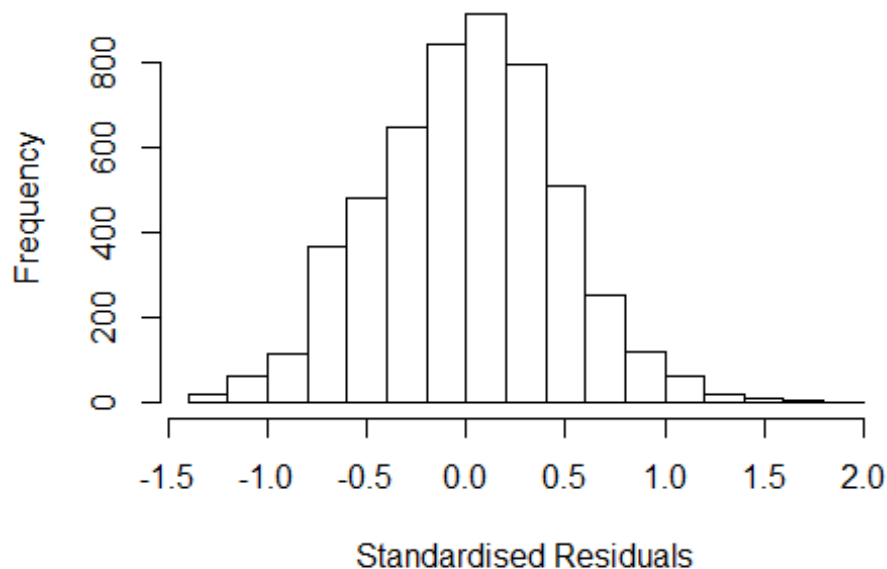


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 2.78354159991838"
## [1] "Male first author team size 2018 geometric mean: 2.19556144001667"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34000, p-value = 4e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.5971761261176"
## [1] "Male last author team size 2018 geometric mean: 2.53797678448564"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.156 1 1.075
## LastAuthorFemale 1.128 1 1.062
## UniqueAuthors 1.146 4 1.017
## Year 1.143 16 1.004
```

## Residuals from first and last author and team size



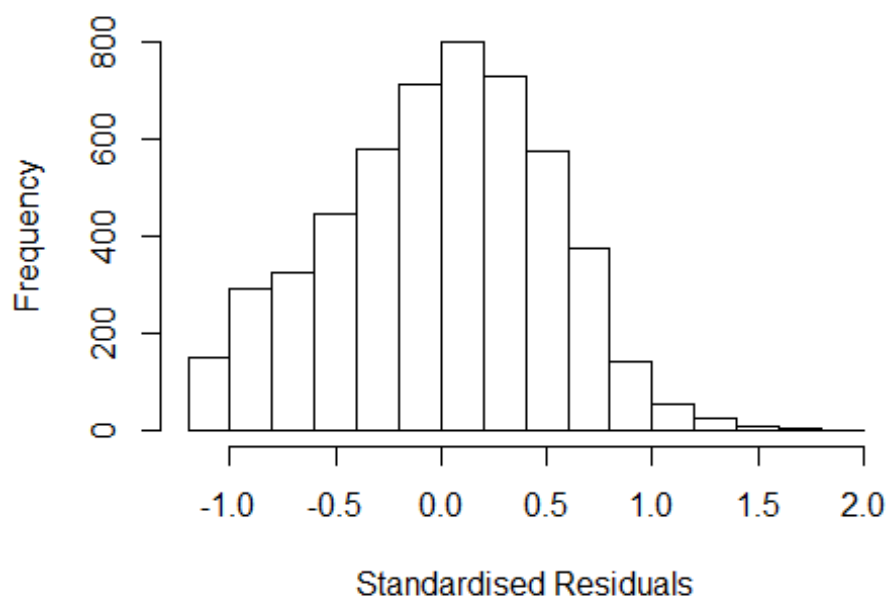
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3490 -0.3114 0.0161 0.3087 1.8795
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.74817 0.03960 18.89 < 2e-16 ***
## FirstAuthorFemale1 -0.02468 0.01404 -1.76 0.0788 .
## LastAuthorFemale1 -0.02969 0.01401 -2.12 0.0342 *
## UniqueAuthors2 0.37219 0.01932 19.27 < 2e-16 ***
## UniqueAuthors3 0.52266 0.01958 26.69 < 2e-16 ***
## UniqueAuthors4 0.57755 0.02166 26.67 < 2e-16 ***
## UniqueAuthors5 0.65048 0.02048 31.76 < 2e-16 ***
## Year1997 -0.03464 0.05163 -0.67 0.5024
## Year1998 0.01738 0.05479 0.32 0.7511
## Year1999 -0.06883 0.05056 -1.36 0.1734
```

```

## Year2000      -0.05181    0.05232   -0.99    0.3221
## Year2001      0.03778    0.05034    0.75    0.4530
## Year2002     -0.00502    0.04790   -0.10    0.9166
## Year2003     -0.00726    0.04954   -0.15    0.8835
## Year2004     -0.06256    0.04530   -1.38    0.1674
## Year2005     -0.07418    0.04806   -1.54    0.1228
## Year2006      0.00240    0.04759    0.05    0.9598
## Year2007     -0.04967    0.04382   -1.13    0.2571
## Year2008     -0.04354    0.04429   -0.98    0.3255
## Year2009     -0.08089    0.04394   -1.84    0.0657 .
## Year2010     -0.10961    0.04400   -2.49    0.0128 *
## Year2011     -0.13663    0.04358   -3.13    0.0017 **
## Year2012     -0.18867    0.04480   -4.21    2.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.456
## Multiple R-squared:  0.233, Adjusted R-squared:  0.23
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 441 weights are ~= 1. The remaining 4767 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0518 0.8690 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.92e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.087 1 1.043
## LastAuthorFemale 1.073 1 1.036
## Year 1.032 16 1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1496 -0.3652 0.0266 0.3706 1.8406
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9768 0.0441 22.15 < 2e-16 ***
## FirstAuthorFemale1 0.0369 0.0156 2.36 0.01841 *
## LastAuthorFemale1 -0.0573 0.0157 -3.64 0.00027 ***
## Year1997 0.0286 0.0587 0.49 0.62628
## Year1998 0.0742 0.0629 1.18 0.23789
## Year1999 -0.0410 0.0580 -0.71 0.47891
## Year2000 0.0306 0.0592 0.52 0.60519
## Year2001 0.1360 0.0570 2.39 0.01703 *
## Year2002 0.1198 0.0535 2.24 0.02518 *
## Year2003 0.1038 0.0561 1.85 0.06439 .
## Year2004 0.0654 0.0520 1.26 0.20826
## Year2005 0.0385 0.0539 0.71 0.47531
```

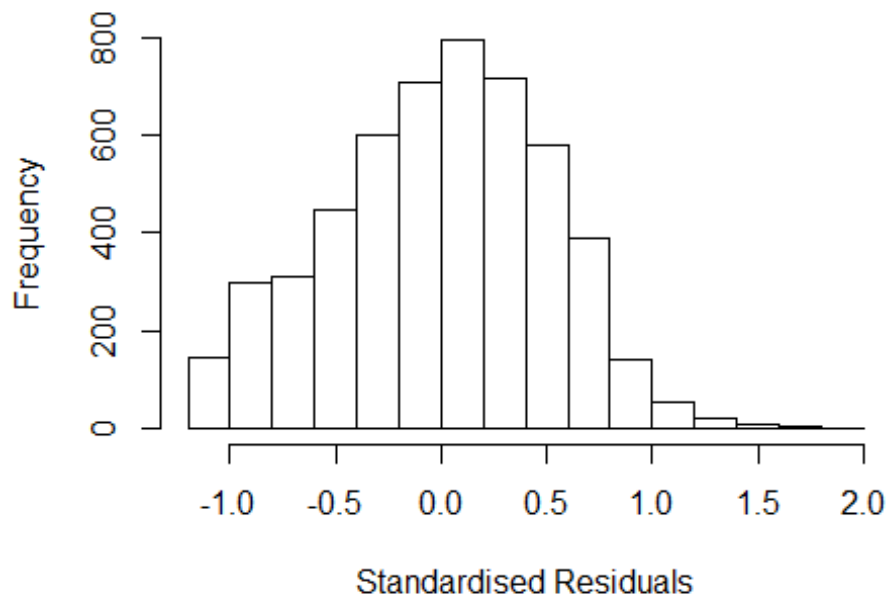


```

## Year2006          0.1191      0.0522      2.28  0.02254 *
## Year2007          0.1002      0.0495      2.02  0.04304 *
## Year2008          0.1177      0.0499      2.36  0.01829 *
## Year2009          0.0257      0.0506      0.51  0.61075
## Year2010         -0.0195      0.0504     -0.39  0.69950
## Year2011         -0.0240      0.0504     -0.48  0.63363
## Year2012         -0.0828      0.0514     -1.61  0.10765
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.0153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 412 weights are ~= 1. The remaining 4796 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.206  0.872  0.949  0.913  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## Year              1.022 16      1.001

```

## Residuals from first author



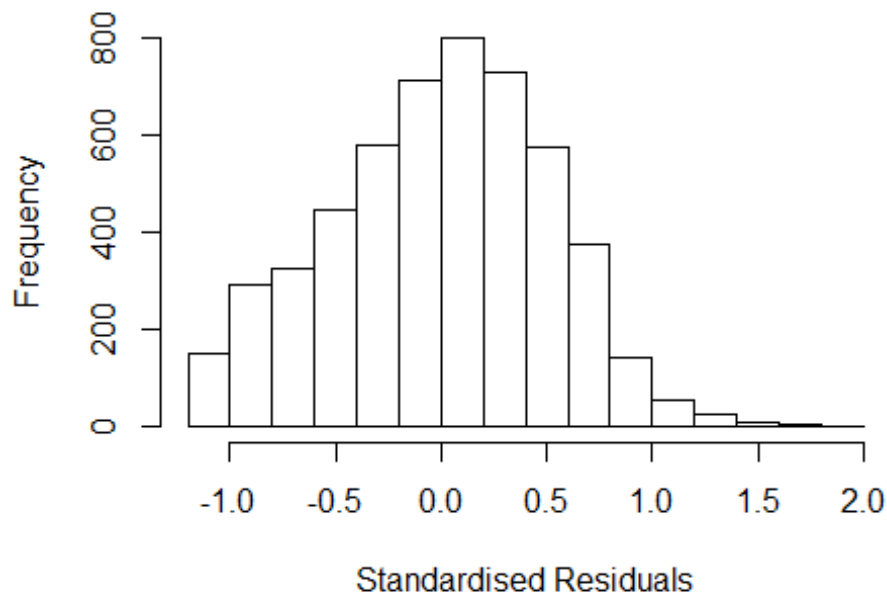
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1185 -0.3611  0.0252  0.3678  1.8541
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9593    0.0437   21.98  <2e-16 ***
## FirstAuthorFemale1 0.0190    0.0152    1.25   0.211
## Year1997        0.0335    0.0584    0.57   0.566
## Year1998        0.0777    0.0628    1.24   0.216
## Year1999       -0.0380    0.0580   -0.65   0.513
## Year2000        0.0345    0.0589    0.59   0.558
## Year2001        0.1401    0.0570    2.46   0.014 *
## Year2002        0.1204    0.0534    2.25   0.024 *
## Year2003        0.1066    0.0560    1.90   0.057 .
## Year2004        0.0688    0.0519    1.33   0.185
## Year2005        0.0388    0.0539    0.72   0.472
## Year2006        0.1244    0.0520    2.39   0.017 *
```

```

## Year2007          0.1019      0.0495      2.06      0.039 *
## Year2008          0.1184      0.0498      2.38      0.017 *
## Year2009          0.0283      0.0506      0.56      0.576
## Year2010         -0.0175      0.0504     -0.35      0.728
## Year2011         -0.0221      0.0504     -0.44      0.660
## Year2012         -0.0817      0.0514     -1.59      0.112
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0162, Adjusted R-squared:  0.0129
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 416 weights are ~= 1. The remaining 4792 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.199  0.872  0.949   0.912  0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.009 1          1.005
## Year            1.009 16          1.000

```

## Residuals from last author



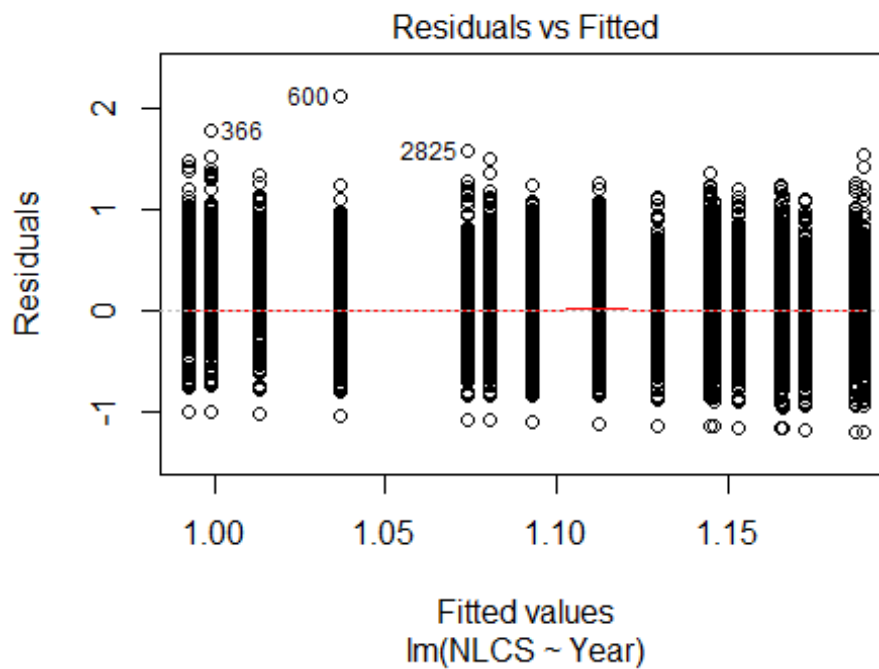
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1266 -0.3614 0.0302 0.3723 1.8283
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9851 0.0436 22.62 <2e-16 ***
## LastAuthorFemale1 -0.0456 0.0153 -2.99 0.0028 **
## Year1997 0.0334 0.0584 0.57 0.5676
## Year1998 0.0764 0.0626 1.22 0.2229
## Year1999 -0.0365 0.0577 -0.63 0.5266
## Year2000 0.0345 0.0588 0.59 0.5568
## Year2001 0.1414 0.0567 2.49 0.0127 *
## Year2002 0.1248 0.0532 2.35 0.0190 *
## Year2003 0.1076 0.0557 1.93 0.0535 .
## Year2004 0.0725 0.0517 1.40 0.1608
## Year2005 0.0418 0.0537 0.78 0.4363
## Year2006 0.1253 0.0518 2.42 0.0156 *
```

```

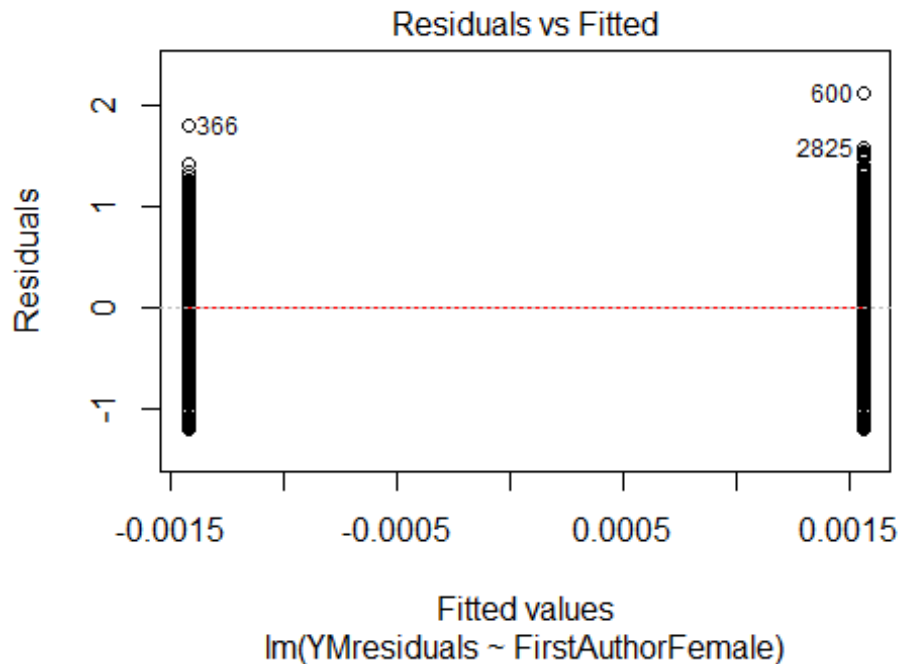
## Year2007          0.1066      0.0492      2.17      0.0303 *
## Year2008          0.1226      0.0495      2.48      0.0133 *
## Year2009          0.0319      0.0503      0.64      0.5252
## Year2010         -0.0136      0.0502     -0.27      0.7859
## Year2011         -0.0152      0.0500     -0.30      0.7611
## Year2012         -0.0741      0.0511     -1.45      0.1474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.532
## Multiple R-squared:  0.0177, Adjusted R-squared:  0.0144
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 416 weights are ~= 1. The remaining 4792 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.213  0.873  0.949  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5208"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3204"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 406 406 404 440 481 444 452 414 457 421 560 609 621 670 708
## 2011 2012
## 762 692
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 311 298 306 352 349 284 366 345 393 356 473 502 527 569 588
## 2011 2012

```

```
## 635 584
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 302 284 279 335 327 264 330 304 344 311 429 454 478 507 547
## 2011 2012
## 591 544
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 200, df = 16, p-value <2e-16
```

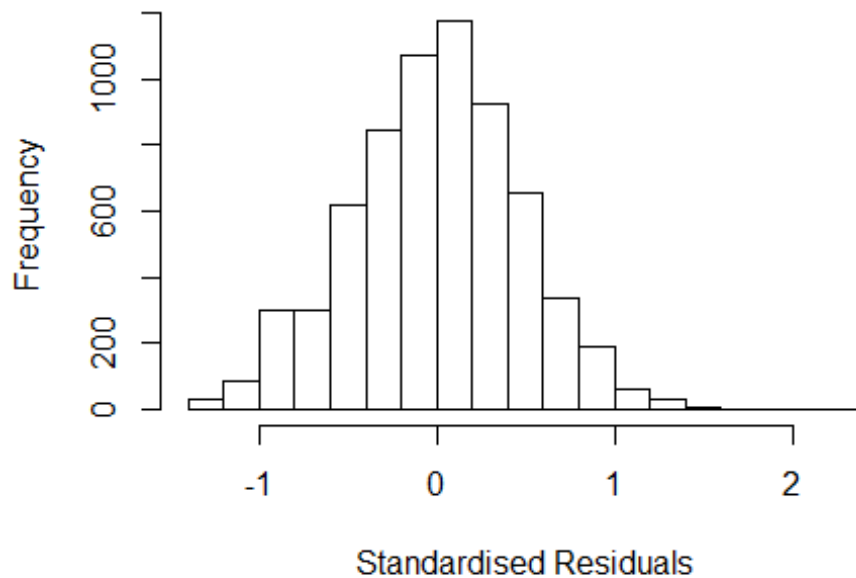


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.7, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 2.77037470692304"
## [1] "Male first author team size 2018 geometric mean: 2.53042502130054"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.62768472273312"
## [1] "Male last author team size 2018 geometric mean: 2.76505354362668"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.128 1      1.062
## LastAuthorFemale  1.091 1      1.044
## UniqueAuthors    1.072 4      1.009
## Year             1.118 16      1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3622 -0.3147 0.0112 0.3132 2.2518
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.86183 0.03923 21.97 < 2e-16 ***
## FirstAuthorFemale1 -0.02852 0.01262 -2.26 0.02392 *
## LastAuthorFemale1 -0.01685 0.01242 -1.36 0.17493
## UniqueAuthors2 0.21435 0.01598 13.42 < 2e-16 ***
## UniqueAuthors3 0.29462 0.01747 16.87 < 2e-16 ***
## UniqueAuthors4 0.34441 0.02166 15.90 < 2e-16 ***
## UniqueAuthors5 0.44019 0.02209 19.93 < 2e-16 ***
## Year1997 0.04938 0.05233 0.94 0.34544
## Year1998 -0.01831 0.05276 -0.35 0.72857
## Year1999 0.00934 0.04943 0.19 0.85008
```

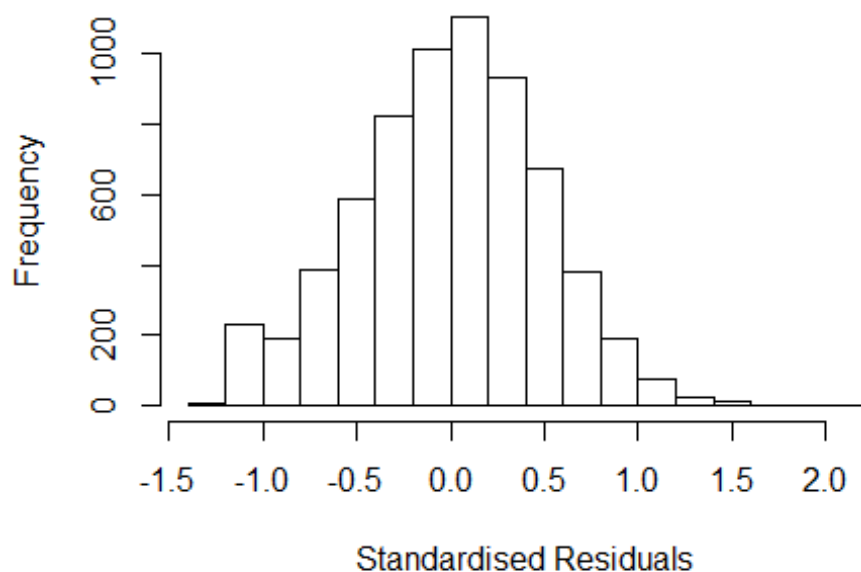


```

## Year2000          0.07529      0.04917      1.53  0.12577
## Year2001          0.06418      0.05067      1.27  0.20539
## Year2002          0.17383      0.04632      3.75  0.00018 ***
## Year2003          0.12245      0.04702      2.60  0.00924 **
## Year2004          0.13187      0.04497      2.93  0.00338 **
## Year2005          0.13103      0.04447      2.95  0.00323 **
## Year2006          0.12989      0.04322      3.01  0.00266 **
## Year2007          0.09685      0.04291      2.26  0.02405 *
## Year2008          0.13052      0.04242      3.08  0.00210 **
## Year2009          0.09937      0.04376      2.27  0.02320 *
## Year2010          0.08363      0.04365      1.92  0.05545 .
## Year2011          0.05449      0.04307      1.27  0.20591
## Year2012          0.08873      0.04438      2.00  0.04562 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.464
## Multiple R-squared:  0.103, Adjusted R-squared:  0.0998
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## observation 372 is an outlier with |weight| = 0 ( < 1.5e-05);
## 545 weights are ~ = 1. The remaining 6084 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.184  0.868  0.950  0.904  0.986  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          1.51e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##           500         50         2         1         1000         200
##   trace.lev    mts    compute.rd
##           0         1000         0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.109 1          1.053
## LastAuthorFemale  1.086 1          1.042
## Year              1.057 16          1.002

```

## Residuals from first and last author



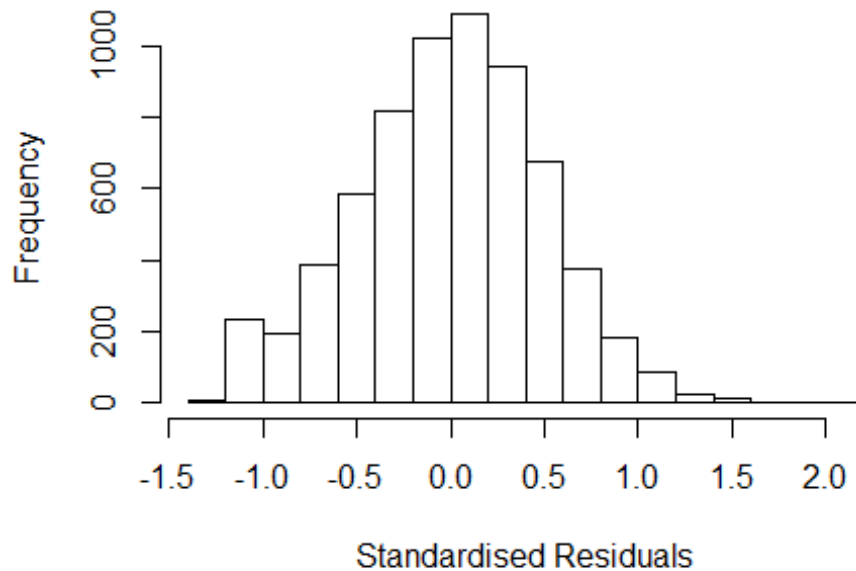
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2135 -0.3261 0.0136 0.3265 2.1277
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.983715 0.041556 23.67 < 2e-16 ***
## FirstAuthorFemale1 0.000345 0.013109 0.03 0.97903
## LastAuthorFemale1 -0.023149 0.012988 -1.78 0.07475 .
## Year1997 0.051547 0.055718 0.93 0.35493
## Year1998 -0.006805 0.056027 -0.12 0.90333
## Year1999 0.026850 0.052370 0.51 0.60817
## Year2000 0.095906 0.052687 1.82 0.06876 .
## Year2001 0.103925 0.054183 1.92 0.05515 .
## Year2002 0.229476 0.049924 4.60 4.4e-06 ***
## Year2003 0.184897 0.050224 3.68 0.00023 ***
## Year2004 0.189380 0.048523 3.90 9.6e-05 ***
## Year2005 0.203420 0.047596 4.27 1.9e-05 ***
```

```

## Year2006          0.206045    0.046136    4.47  8.1e-06 ***
## Year2007          0.167654    0.045651    3.67  0.00024 ***
## Year2008          0.216323    0.045143    4.79  1.7e-06 ***
## Year2009          0.176174    0.046287    3.81  0.00014 ***
## Year2010          0.165212    0.046602    3.55  0.00040 ***
## Year2011          0.136713    0.045671    2.99  0.00277 **
## Year2012          0.178236    0.047111    3.78  0.00016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0189, Adjusted R-squared:  0.0162
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 571 weights are ~= 1. The remaining 6059 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0175 0.8690 0.9510 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## Year      1.042 16      1.001

```

## Residuals from first author



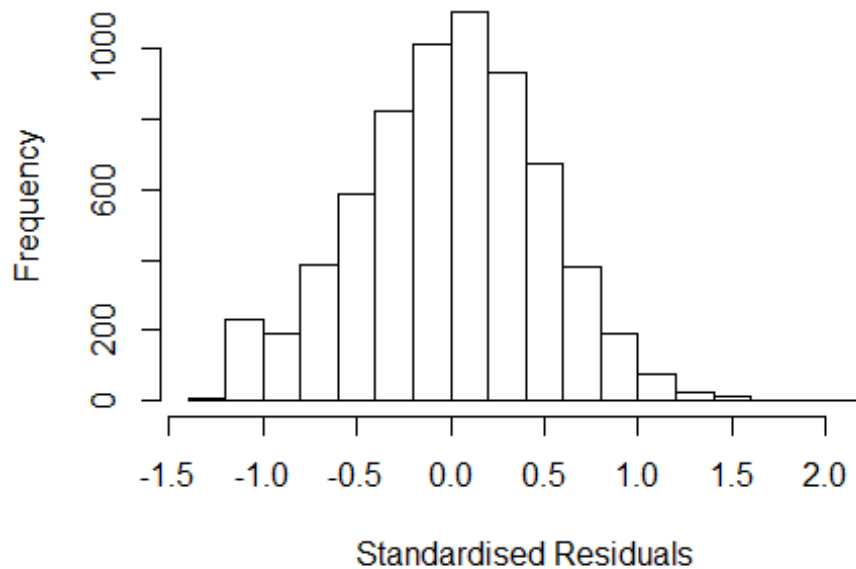
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2070 -0.3262 0.0138 0.3274 2.1325
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.97724 0.04140 23.60 < 2e-16 ***
## FirstAuthorFemale1 -0.00752 0.01275 -0.59 0.55541
## Year1997 0.05329 0.05571 0.96 0.33882
## Year1998 -0.00490 0.05608 -0.09 0.93039
## Year1999 0.02654 0.05247 0.51 0.61302
## Year2000 0.09542 0.05267 1.81 0.07009 .
## Year2001 0.10488 0.05414 1.94 0.05276 .
## Year2002 0.22973 0.04993 4.60 4.3e-06 ***
## Year2003 0.18448 0.05023 3.67 0.00024 ***
## Year2004 0.18926 0.04855 3.90 9.8e-05 ***
## Year2005 0.20369 0.04761 4.28 1.9e-05 ***
## Year2006 0.20682 0.04616 4.48 7.6e-06 ***
```

```

## Year2007          0.16722      0.04567      3.66  0.00025 ***
## Year2008          0.21633      0.04516      4.79  1.7e-06 ***
## Year2009          0.17584      0.04631      3.80  0.00015 ***
## Year2010          0.16549      0.04664      3.55  0.00039 ***
## Year2011          0.13565      0.04569      2.97  0.00300 **
## Year2012          0.17737      0.04715      3.76  0.00017 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.487
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.0159
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 560 weights are ~= 1. The remaining 6070 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0164 0.8700 0.9510 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.021 1      1.010
## Year              1.021 16      1.001

```

## Residuals from last author



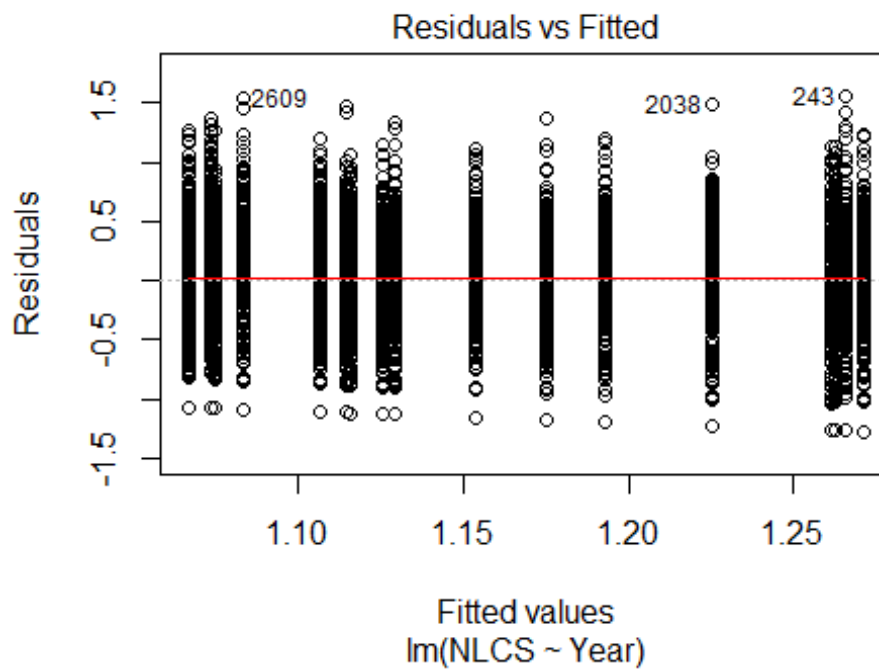
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2133 -0.3260  0.0136  0.3265  2.1276
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.98383    0.04137   23.78 < 2e-16 ***
## LastAuthorFemale1 -0.02303    0.01264   -1.82  0.06855 .
## Year1997         0.05152    0.05571    0.92  0.35509
## Year1998        -0.00681    0.05602   -0.12  0.90327
## Year1999         0.02682    0.05235    0.51  0.60840
## Year2000         0.09590    0.05266    1.82  0.06861 .
## Year2001         0.10392    0.05417    1.92  0.05512 .
## Year2002         0.22948    0.04990    4.60  4.3e-06 ***
## Year2003         0.18488    0.05022    3.68  0.00023 ***
## Year2004         0.18938    0.04851    3.90  9.6e-05 ***
## Year2005         0.20345    0.04753    4.28  1.9e-05 ***
## Year2006         0.20606    0.04609    4.47  7.9e-06 ***
```

```

## Year2007      0.16767      0.04562      3.68  0.00024 ***
## Year2008      0.21635      0.04510      4.80  1.6e-06 ***
## Year2009      0.17618      0.04626      3.81  0.00014 ***
## Year2010      0.16524      0.04650      3.55  0.00038 ***
## Year2011      0.13674      0.04558      3.00  0.00271 **
## Year2012      0.17826      0.04706      3.79  0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.488
## Multiple R-squared:  0.0189, Adjusted R-squared:  0.0164
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 572 weights are ~= 1. The remaining 6058 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0176 0.8690 0.9510 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.51e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6630"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3205"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 409 400 312 314 367 396 393 373 373 438 506 550 605 649 618
## 2011 2012
## 641 601
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 240 232 241 237 261 231 308 298 289 350 401 452 484 554 511
## 2011 2012

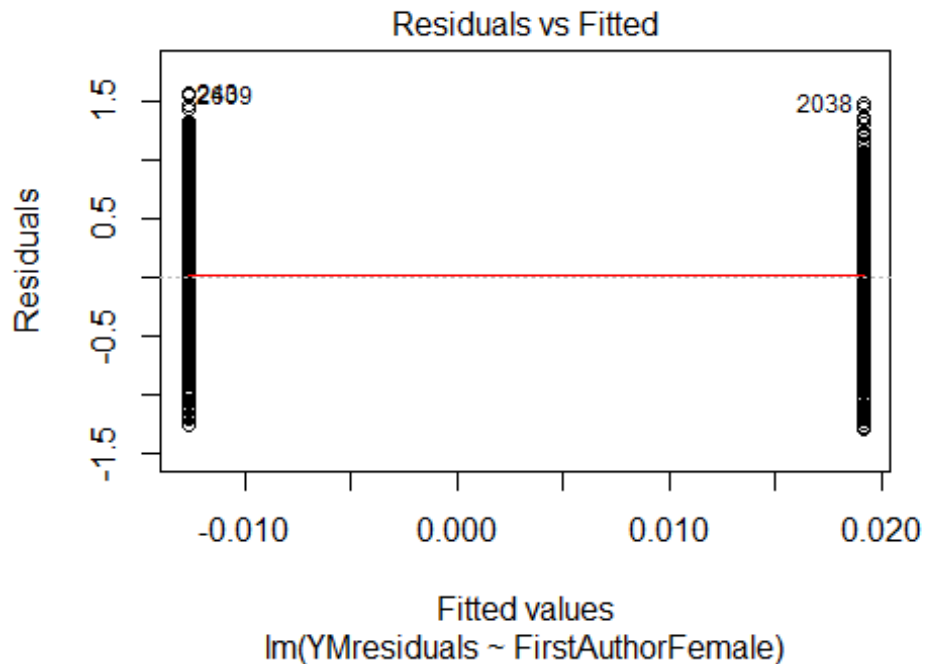
```

```
## 540 489
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 225 213 219 217 240 199 278 275 252 316 364 397 422 488 461
## 2011 2012
## 490 443
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 150, df = 16, p-value <2e-16
```



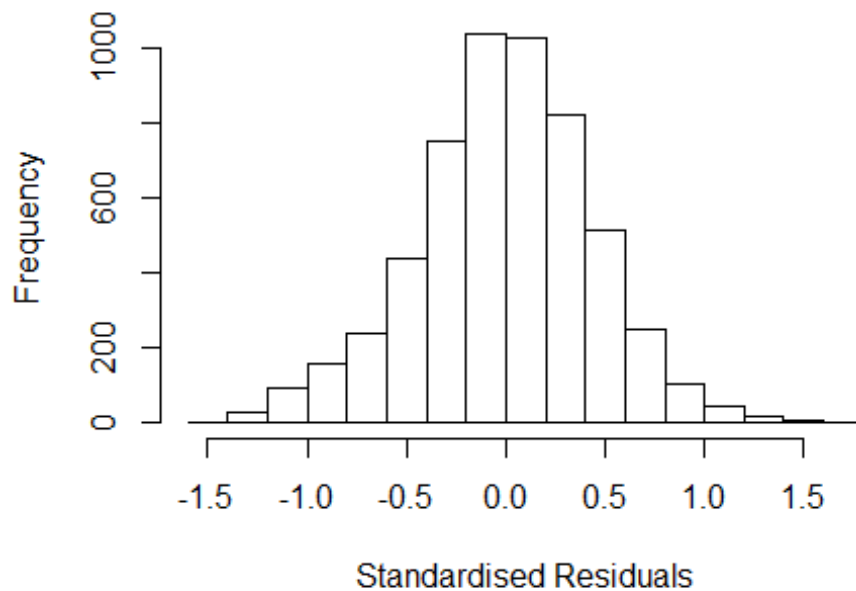
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 21, df = 1, p-value = 5e-06
```





```
## [1] "Female first author team size 2018 geometric mean: 2.83686686395782"
## [1] "Male first author team size 2018 geometric mean: 2.4760095620011"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.70260861581614"
## [1] "Male last author team size 2018 geometric mean: 2.62502532154692"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 28000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1      1.021
## LastAuthorFemale  1.037 1      1.018
## UniqueAuthors    1.091 4      1.011
## Year             1.108 16      1.003
```

## Residuals from first and last author and team size



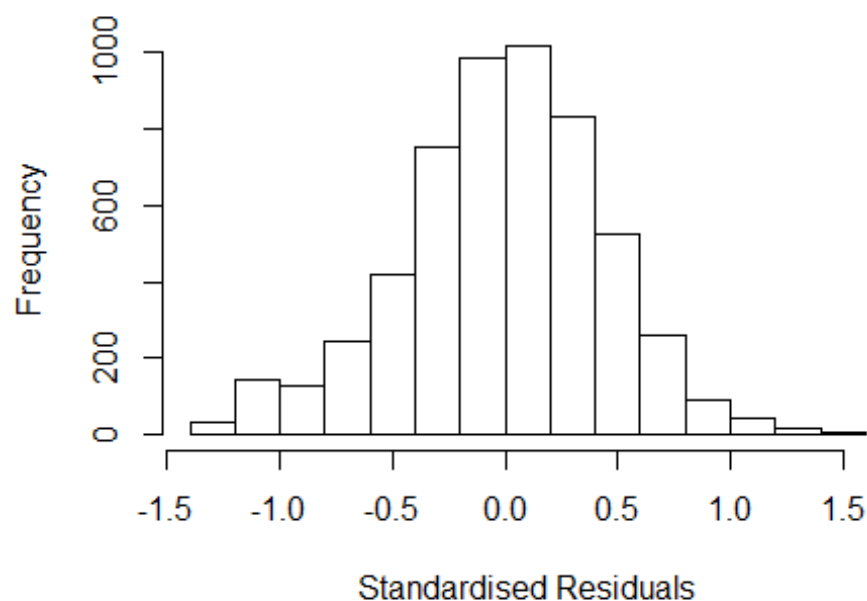
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.40232 -0.27448 0.00291 0.27672 1.69962
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12138 0.03534 31.73 < 2e-16 ***
## FirstAuthorFemale1 0.01446 0.01214 1.19 0.2335
## LastAuthorFemale1 -0.01130 0.01268 -0.89 0.3725
## UniqueAuthors2 0.17752 0.01934 9.18 < 2e-16 ***
## UniqueAuthors3 0.21992 0.02001 10.99 < 2e-16 ***
## UniqueAuthors4 0.25717 0.02317 11.10 < 2e-16 ***
## UniqueAuthors5 0.31208 0.02370 13.17 < 2e-16 ***
## Year1997 -0.00293 0.05077 -0.06 0.9541
## Year1998 -0.00027 0.04717 -0.01 0.9954
## Year1999 -0.02221 0.04549 -0.49 0.6254
```

```

## Year2000      -0.03429    0.04414   -0.78    0.4372
## Year2001      -0.15672    0.05533   -2.83    0.0046 **
## Year2002      -0.12040    0.04280   -2.81    0.0049 **
## Year2003      -0.12394    0.04149   -2.99    0.0028 **
## Year2004      -0.10565    0.04091   -2.58    0.0098 **
## Year2005      -0.17054    0.03936   -4.33    1.5e-05 ***
## Year2006      -0.17722    0.03903   -4.54    5.7e-06 ***
## Year2007      -0.18084    0.03714   -4.87    1.1e-06 ***
## Year2008      -0.17460    0.03801   -4.59    4.4e-06 ***
## Year2009      -0.18458    0.03744   -4.93    8.4e-07 ***
## Year2010      -0.20141    0.03808   -5.29    1.3e-07 ***
## Year2011      -0.24463    0.03854   -6.35    2.4e-10 ***
## Year2012      -0.24743    0.04097   -6.04    1.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.416
## Multiple R-squared:  0.0697, Adjusted R-squared:  0.0659
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 501 weights are ~= 1. The remaining 4998 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0568 0.8610 0.9520 0.8960 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.82e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1 1.011
## LastAuthorFemale 1.032 1 1.016
## Year 1.039 16 1.001

```

## Residuals from first and last author



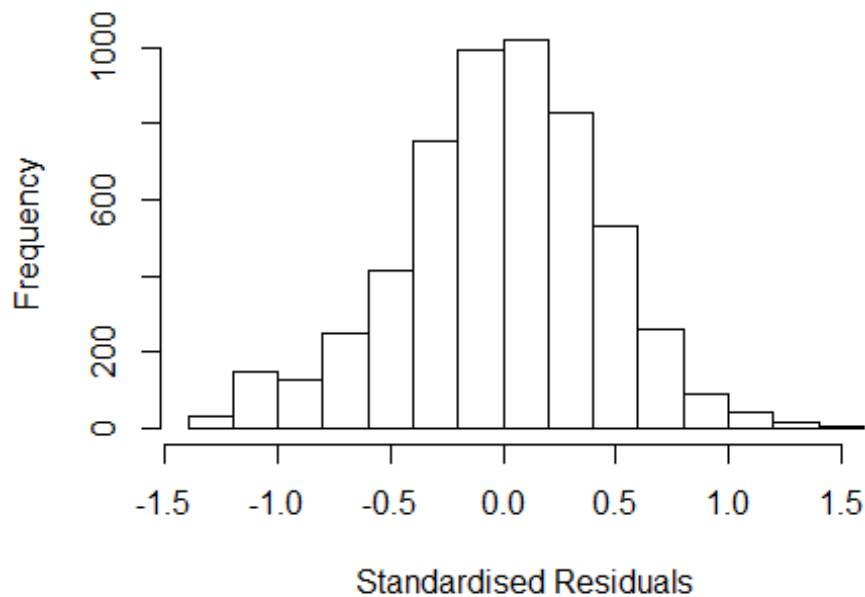
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29565 -0.28172  0.00854  0.28912  1.55918
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.261819   0.034143   36.96 < 2e-16 ***
## FirstAuthorFemale1  0.035625   0.012258    2.91  0.00367 **
## LastAuthorFemale1 -0.008268   0.012885   -0.64  0.52110
## Year1997         0.006478   0.052330    0.12  0.90149
## Year1998         0.000752   0.049180    0.02  0.98780
## Year1999        -0.004187   0.047011   -0.09  0.92904
## Year2000        -0.031077   0.046146   -0.67  0.50069
## Year2001        -0.143262   0.057682   -2.48  0.01303 *
## Year2002        -0.091688   0.043900   -2.09  0.03679 *
## Year2003        -0.086457   0.042467   -2.04  0.04181 *
## Year2004        -0.083929   0.042250   -1.99  0.04703 *
## Year2005        -0.130960   0.040850   -3.21  0.00135 **
```

```

## Year2006      -0.148480    0.039983    -3.71    0.00021 ***
## Year2007      -0.146358    0.038187    -3.83    0.00013 ***
## Year2008      -0.147268    0.038883    -3.79    0.00015 ***
## Year2009      -0.146946    0.038219    -3.84    0.00012 ***
## Year2010      -0.175598    0.039429    -4.45    8.6e-06 ***
## Year2011      -0.210651    0.039496    -5.33    1.0e-07 ***
## Year2012      -0.209305    0.041849    -5.00    5.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0238, Adjusted R-squared:  0.0206
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 475 weights are ~= 1. The remaining 5024 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.144  0.862  0.950  0.896  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.013 1      1.007
## Year      1.013 16      1.000

```

## Residuals from first author



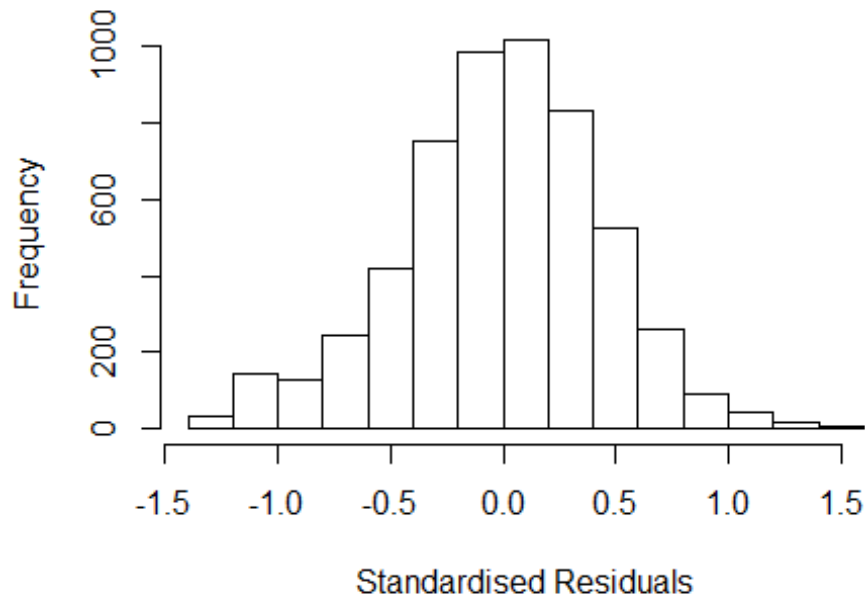
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.30109 -0.28181 0.00928 0.28855 1.56144
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25956 0.03385 37.21 < 2e-16 ***
## FirstAuthorFemale1 0.03419 0.01226 2.79 0.00532 **
## Year1997 0.00735 0.05222 0.14 0.88814
## Year1998 0.00136 0.04915 0.03 0.97786
## Year1999 -0.00339 0.04694 -0.07 0.94246
## Year2000 -0.03064 0.04610 -0.66 0.50632
## Year2001 -0.14307 0.05759 -2.48 0.01301 *
## Year2002 -0.09136 0.04386 -2.08 0.03730 *
## Year2003 -0.08666 0.04244 -2.04 0.04124 *
## Year2004 -0.08396 0.04224 -1.99 0.04688 *
## Year2005 -0.13123 0.04083 -3.21 0.00132 **
## Year2006 -0.14836 0.03997 -3.71 0.00021 ***
```

```

## Year2007          -0.14658      0.03817    -3.84  0.00012 ***
## Year2008          -0.14748      0.03886    -3.80  0.00015 ***
## Year2009          -0.14684      0.03819    -3.84  0.00012 ***
## Year2010          -0.17544      0.03940    -4.45  8.7e-06 ***
## Year2011          -0.21048      0.03946    -5.33  1.0e-07 ***
## Year2012          -0.21011      0.04181    -5.03  5.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0238, Adjusted R-squared:  0.0207
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 483 weights are ~= 1. The remaining 5016 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.143  0.862  0.949  0.896  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.011
## Year              1.023 16          1.001

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27918 -0.28588 0.00884 0.28858 1.55124
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26976 0.03390 37.46 < 2e-16 ***
## LastAuthorFemale1 -0.00163 0.01288 -0.13 0.89929
## Year1997 0.00942 0.05244 0.18 0.85746
## Year1998 0.00386 0.04908 0.08 0.93734
## Year1999 -0.00323 0.04704 -0.07 0.94520
## Year2000 -0.02936 0.04613 -0.64 0.52447
## Year2001 -0.14055 0.05780 -2.43 0.01506 *
## Year2002 -0.08732 0.04382 -1.99 0.04633 *
## Year2003 -0.08493 0.04241 -2.00 0.04527 *
## Year2004 -0.07900 0.04223 -1.87 0.06142 .
## Year2005 -0.12672 0.04090 -3.10 0.00196 **
## Year2006 -0.14432 0.03995 -3.61 0.00031 ***
```

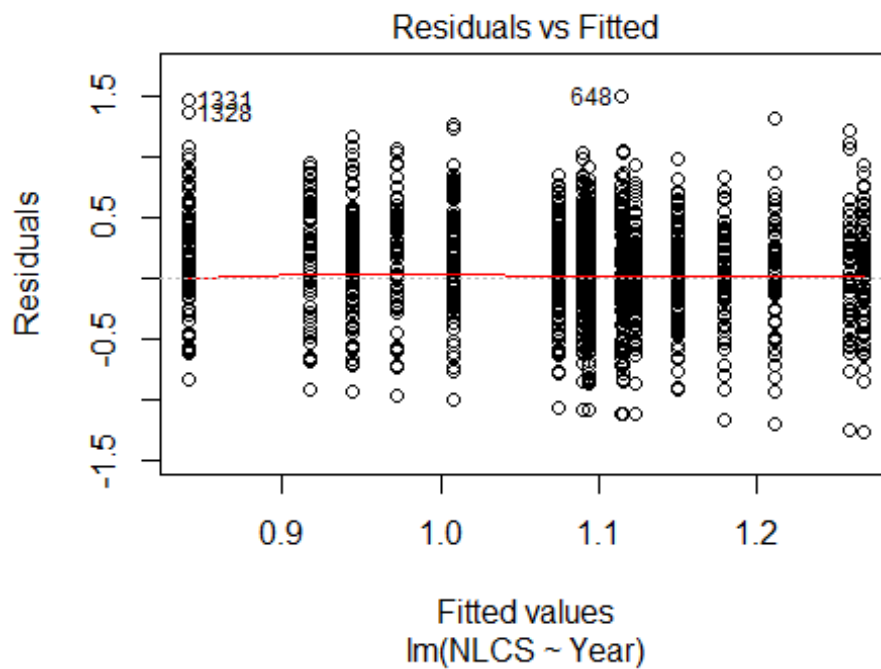


```

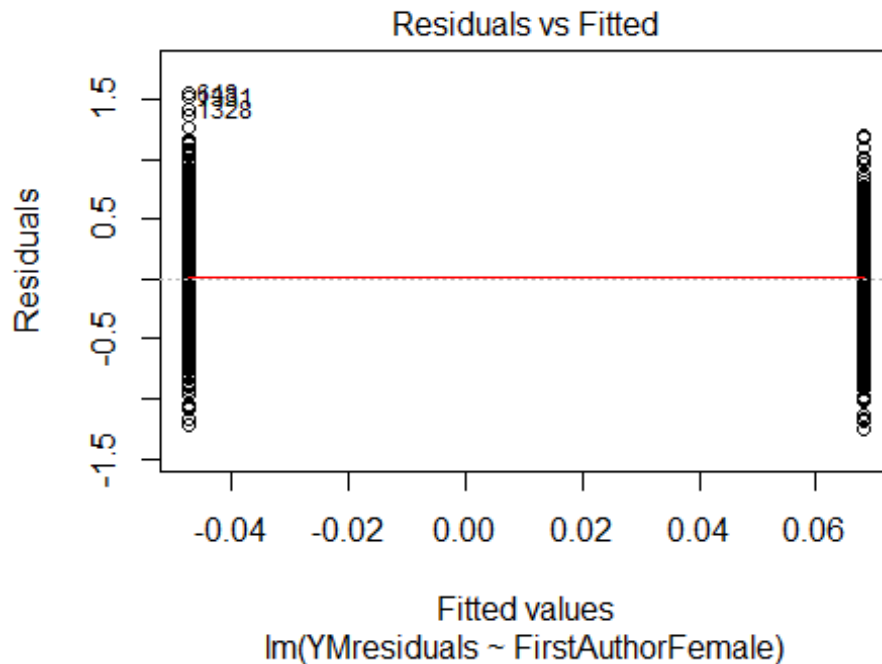
## Year2007          -0.14239      0.03813    -3.73  0.00019 ***
## Year2008          -0.14250      0.03888    -3.66  0.00025 ***
## Year2009          -0.14160      0.03815    -3.71  0.00021 ***
## Year2010          -0.16907      0.03938    -4.29  1.8e-05 ***
## Year2011          -0.20388      0.03943    -5.17  2.4e-07 ***
## Year2012          -0.20288      0.04175    -4.86  1.2e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.423
## Multiple R-squared:  0.0225, Adjusted R-squared:  0.0194
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 481 weights are ~= 1. The remaining 5018 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.150  0.862  0.950  0.896  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.82e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5499"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3206"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   93  132  113   82  143  182  155  137  165  163  215  188  201  181  156
## 2011 2012
##  188  173
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   55   79   58   58   56   67  116   97  120  126  172  155  153  144  128
## 2011 2012

```

```
## 151 139
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 49 65 50 52 54 61 108 88 109 117 158 144 133 128 112
## 2011 2012
## 138 123
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 140, df = 16, p-value <2e-16
```

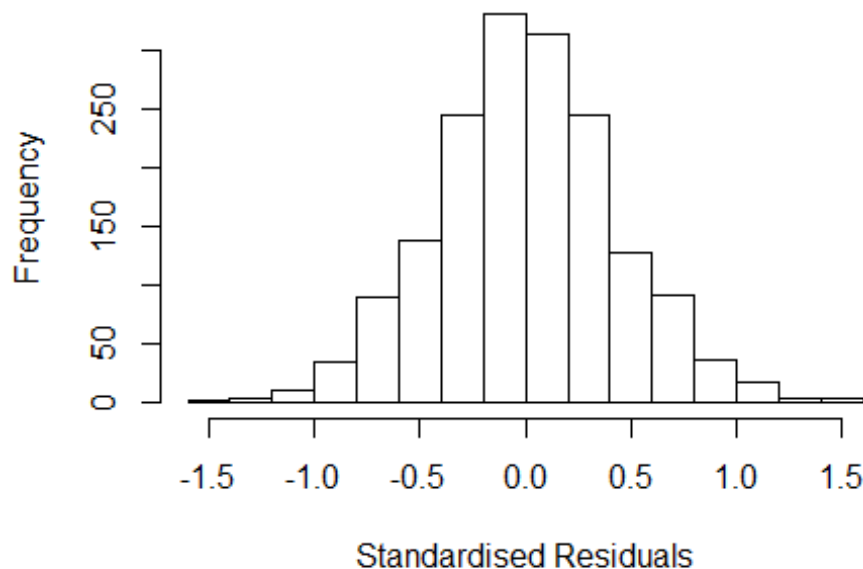


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 24, df = 1, p-value = 1e-06
```



```
## [1] "Female first author team size 2018 geometric mean: 3.06588612247746"
## [1] "Male first author team size 2018 geometric mean: 2.85743110071379"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2500, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.00071230743575"
## [1] "Male last author team size 2018 geometric mean: 2.94862643507021"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.079 1          1.039
## LastAuthorFemale  1.098 1          1.048
## UniqueAuthors    1.317 4          1.035
## Year             1.360 16          1.010
```

## Residuals from first and last author and team size



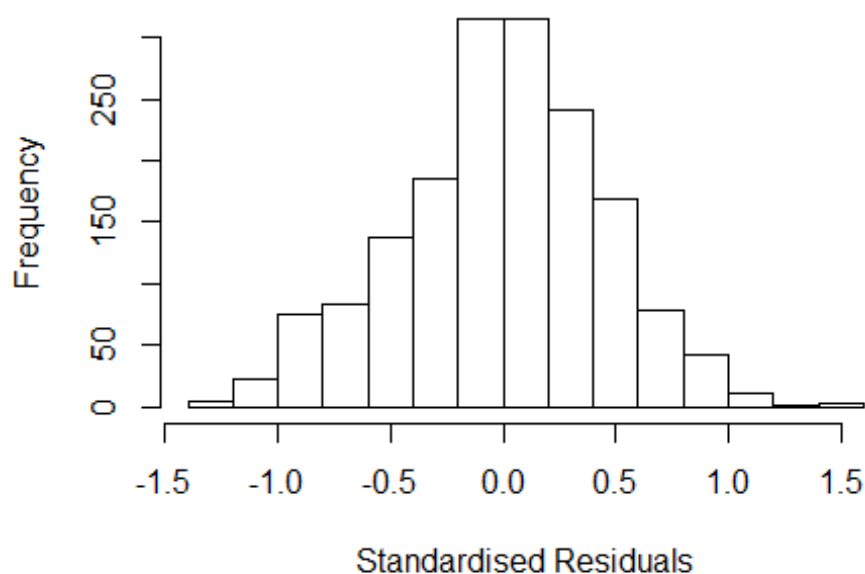
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.45756 -0.26647 -0.00354  0.27177  1.46812
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8786    0.0770   11.41 < 2e-16 ***
## FirstAuthorFemale1 0.0959    0.0211    4.54 6.1e-06 ***
## LastAuthorFemale1 0.0447    0.0224    1.99 0.04633 *
## UniqueAuthors2    0.3328    0.0387    8.60 < 2e-16 ***
## UniqueAuthors3    0.4358    0.0384   11.35 < 2e-16 ***
## UniqueAuthors4    0.4416    0.0428   10.31 < 2e-16 ***
## UniqueAuthors5    0.4686    0.0419   11.19 < 2e-16 ***
## Year1997          0.0386    0.0927    0.42 0.67715
## Year1998          0.0391    0.0960    0.41 0.68382
## Year1999         -0.0383    0.0991   -0.39 0.69904
```

```

## Year2000          -0.0302      0.1076   -0.28   0.77902
## Year2001          -0.2268      0.1090   -2.08   0.03761 *
## Year2002          -0.2139      0.0880   -2.43   0.01522 *
## Year2003          -0.2738      0.0908   -3.01   0.00262 **
## Year2004          -0.3665      0.0872   -4.20   2.8e-05 ***
## Year2005          -0.3007      0.0832   -3.62   0.00031 ***
## Year2006          -0.1522      0.0807   -1.89   0.05938 .
## Year2007          -0.2046      0.0774   -2.64   0.00829 **
## Year2008          -0.1650      0.0791   -2.08   0.03727 *
## Year2009          -0.1680      0.0775   -2.17   0.03024 *
## Year2010          -0.1797      0.0826   -2.18   0.02970 *
## Year2011          -0.1953      0.0796   -2.45   0.01424 *
## Year2012          -0.2380      0.0807   -2.95   0.00323 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.408
## Multiple R-squared:  0.207, Adjusted R-squared:  0.197
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 144 weights are ~= 1. The remaining 1545 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.167  0.856  0.952   0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          5.92e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.081 1 1.040
## LastAuthorFemale 1.064 1 1.032
## Year 1.122 16 1.004

```

## Residuals from first and last author



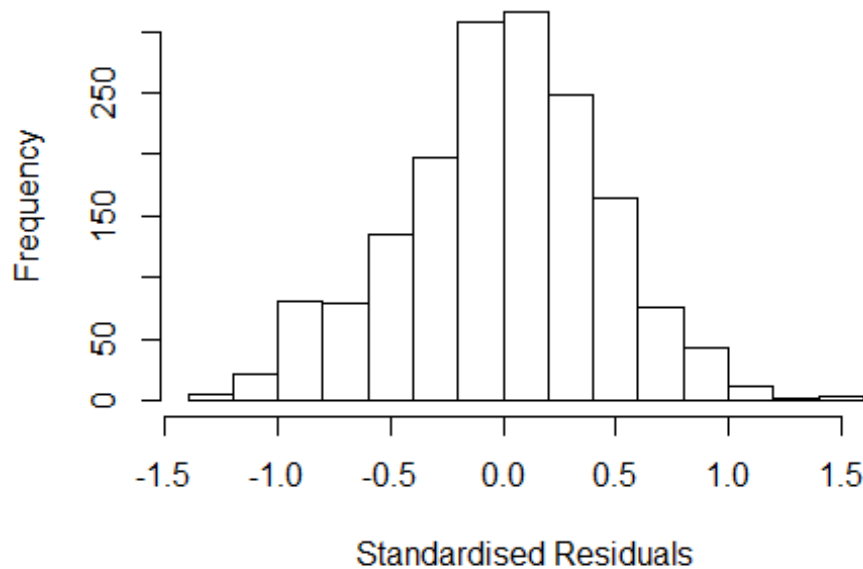
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2964 -0.2977 0.0114 0.3084 1.5823
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1652 0.0771 15.11 < 2e-16 ***
## FirstAuthorFemale1 0.1339 0.0223 6.00 2.5e-09 ***
## LastAuthorFemale1 0.0402 0.0233 1.73 0.0847 .
## Year1997 0.0294 0.0942 0.31 0.7548
## Year1998 0.0549 0.0979 0.56 0.5749
## Year1999 -0.0460 0.0980 -0.47 0.6392
## Year2000 -0.0429 0.1156 -0.37 0.7106
## Year2001 -0.2683 0.1233 -2.18 0.0297 *
## Year2002 -0.1956 0.0996 -1.96 0.0496 *
## Year2003 -0.2924 0.1053 -2.78 0.0056 ***
## Year2004 -0.4345 0.0981 -4.43 1.0e-05 ***
## Year2005 -0.2880 0.0936 -3.08 0.0021 **
```

```

## Year2006          -0.1323      0.0848   -1.56   0.1189
## Year2007          -0.1403      0.0811   -1.73   0.0839 .
## Year2008          -0.1072      0.0819   -1.31   0.1909
## Year2009          -0.0983      0.0810   -1.21   0.2249
## Year2010          -0.0943      0.0861   -1.10   0.2733
## Year2011          -0.1420      0.0826   -1.72   0.0857 .
## Year2012          -0.1886      0.0845   -2.23   0.0258 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.428
## Multiple R-squared:  0.0815, Adjusted R-squared:  0.0716
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 168 weights are ~= 1. The remaining 1521 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.142  0.857  0.942  0.894  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      5.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.071 1          1.035
## Year              1.071 16          1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26566 -0.29345 0.00892 0.30181 1.57281
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1752 0.0771 15.24 < 2e-16 ***
## FirstAuthorFemale1 0.1378 0.0227 6.07 1.6e-09 ***
## Year1997 0.0321 0.0948 0.34 0.7352
## Year1998 0.0594 0.0986 0.60 0.5470
## Year1999 -0.0483 0.0984 -0.49 0.6235
## Year2000 -0.0473 0.1162 -0.41 0.6839
## Year2001 -0.2653 0.1249 -2.12 0.0338 *
## Year2002 -0.1938 0.1002 -1.93 0.0534 .
## Year2003 -0.2902 0.1058 -2.74 0.0062 **
## Year2004 -0.4350 0.0987 -4.41 1.1e-05 ***
## Year2005 -0.2880 0.0944 -3.05 0.0023 **
## Year2006 -0.1301 0.0854 -1.52 0.1281
```

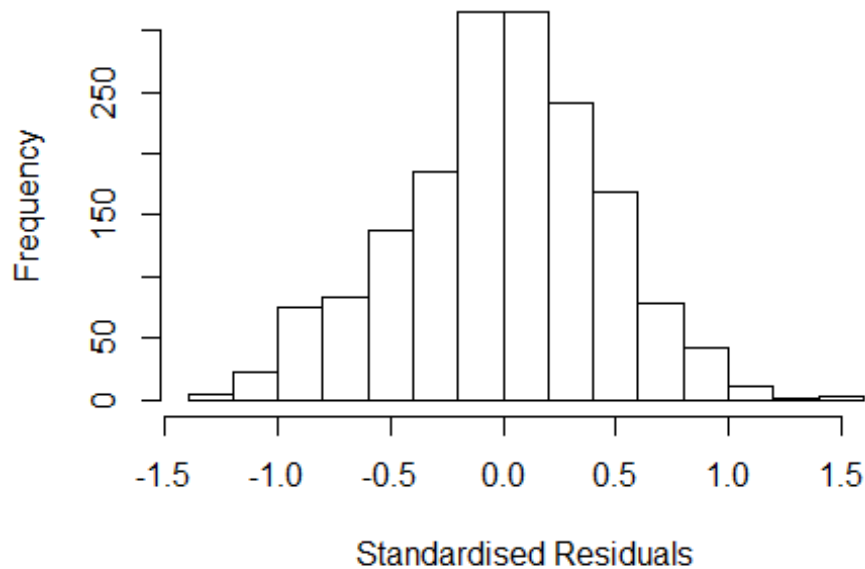


```

## Year2007          -0.1341      0.0817   -1.64   0.1010
## Year2008          -0.1050      0.0824   -1.27   0.2029
## Year2009          -0.0969      0.0816   -1.19   0.2353
## Year2010          -0.0934      0.0866   -1.08   0.2812
## Year2011          -0.1386      0.0832   -1.67   0.0957 .
## Year2012          -0.1834      0.0851   -2.15   0.0314 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0792, Adjusted R-squared:  0.0699
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 161 weights are ~= 1. The remaining 1528 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.150  0.858  0.944  0.894  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.062 1          1.030
## Year            1.062 16          1.002

```

## Residuals from last author



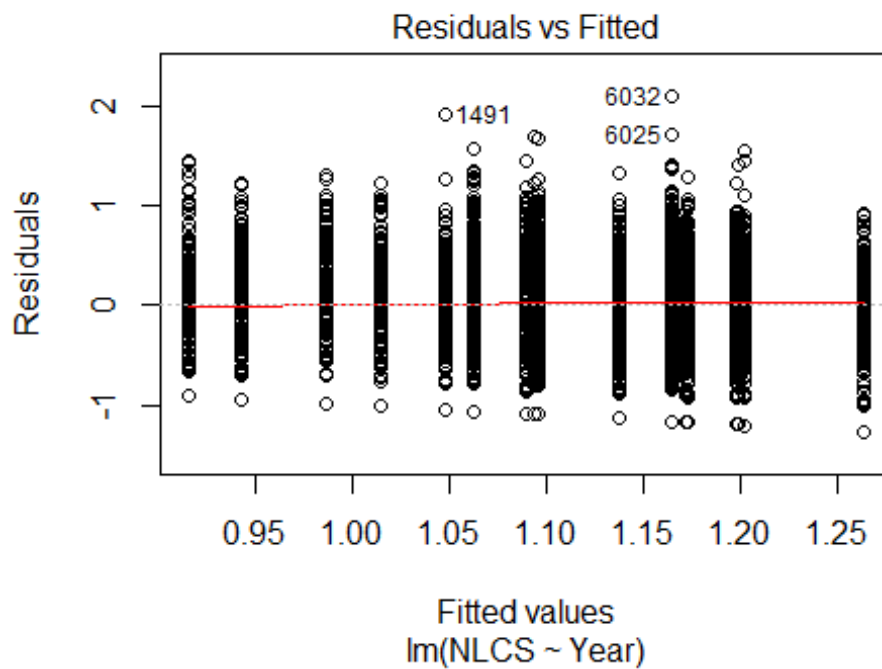
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26661 -0.29394  0.00844  0.30221  1.53913
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2064     0.0745   16.19 < 2e-16 ***
## LastAuthorFemale1  0.0544     0.0239    2.28  0.0230 *
## Year1997          0.0496     0.0923    0.54  0.5909
## Year1998          0.0602     0.0979    0.61  0.5388
## Year1999         -0.0356     0.0951   -0.37  0.7079
## Year2000         -0.0465     0.1146   -0.41  0.6848
## Year2001         -0.2604     0.1243   -2.09  0.0364 *
## Year2002         -0.1922     0.0979   -1.96  0.0497 *
## Year2003         -0.3032     0.1055   -2.87  0.0041 **
## Year2004         -0.4326     0.0980   -4.41 1.1e-05 ***
## Year2005         -0.2703     0.0931   -2.90  0.0037 **
## Year2006         -0.1249     0.0831   -1.50  0.1330
```

```

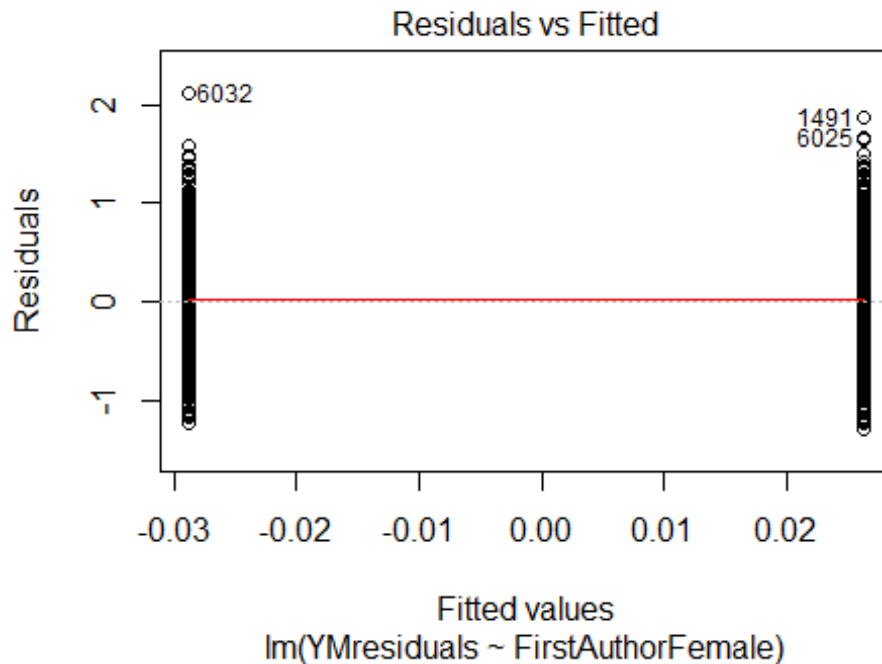
## Year2007          -0.1359      0.0792   -1.72   0.0864 .
## Year2008          -0.0906      0.0803   -1.13   0.2596
## Year2009          -0.0837      0.0790   -1.06   0.2890
## Year2010          -0.0902      0.0839   -1.07   0.2829
## Year2011          -0.1202      0.0806   -1.49   0.1362
## Year2012          -0.1586      0.0825   -1.92   0.0547 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.429
## Multiple R-squared:  0.0615, Adjusted R-squared:  0.052
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 143 weights are ~= 1. The remaining 1546 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.172  0.853  0.946  0.894  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1689"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3207"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  216  200  190  225  232  244  235  206  228  225  251  339  323  411  542
## 2011 2012
##  527  512
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  159  154  143  175  172  178  206  172  197  190  212  285  279  360  467
## 2011 2012

```

```
## 471 433
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 152 148 137 169 159 171 196 161 186 182 196 262 260 338 440
## 2011 2012
## 445 411
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 56, df = 16, p-value = 3e-06
```

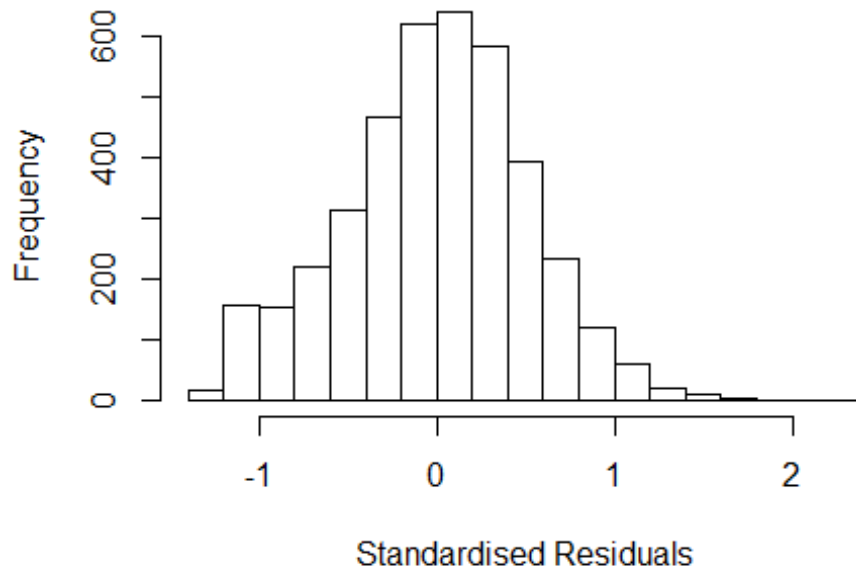


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.027, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 2.1510373236138"
## [1] "Male first author team size 2018 geometric mean: 1.69604622132888"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 4e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.01464190907388"
## [1] "Male last author team size 2018 geometric mean: 1.81170765223814"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 29000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.209 1          1.100
## LastAuthorFemale  1.213 1          1.101
## UniqueAuthors    1.122 4          1.014
## Year              1.149 16         1.004
```

## Residuals from first and last author and team size



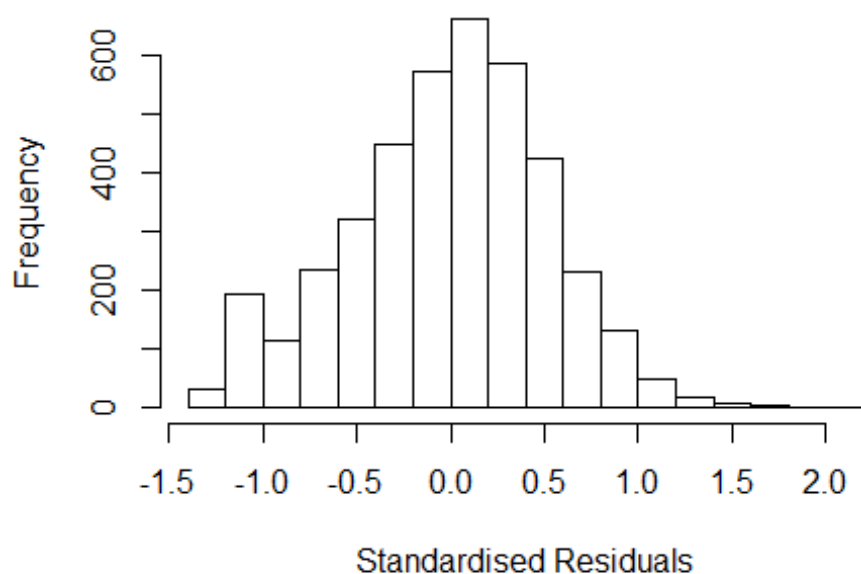
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3038 -0.3343  0.0168  0.3451  2.2025
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8666    0.0505   17.17 < 2e-16 ***
## FirstAuthorFemale1 -0.0564    0.0184   -3.07  0.00216 **
## LastAuthorFemale1 -0.0120    0.0185   -0.65  0.51574
## UniqueAuthors2     0.1283    0.0211    6.07  1.4e-09 ***
## UniqueAuthors3     0.1512    0.0232    6.53  7.6e-11 ***
## UniqueAuthors4     0.1932    0.0325    5.95  3.0e-09 ***
## UniqueAuthors5     0.2522    0.0305    8.26 < 2e-16 ***
## Year1997          0.0343    0.0693    0.49  0.62122
## Year1998          0.1040    0.0699    1.49  0.13683
## Year1999          0.1016    0.0647    1.57  0.11609
```

```

## Year2000          0.1690      0.0682      2.48  0.01326 *
## Year2001          0.1361      0.0625      2.18  0.02966 *
## Year2002          0.2679      0.0595      4.51  6.8e-06 ***
## Year2003          0.2485      0.0616      4.03  5.6e-05 ***
## Year2004          0.3004      0.0628      4.78  1.8e-06 ***
## Year2005          0.2733      0.0625      4.37  1.3e-05 ***
## Year2006          0.3484      0.0593      5.87  4.6e-09 ***
## Year2007          0.1935      0.0572      3.38  0.00073 ***
## Year2008          0.2755      0.0573      4.81  1.6e-06 ***
## Year2009          0.1863      0.0584      3.19  0.00145 **
## Year2010          0.1518      0.0562      2.70  0.00696 **
## Year2011          0.1816      0.0567      3.20  0.00137 **
## Year2012          0.2533      0.0573      4.42  1.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.502
## Multiple R-squared:  0.0556, Adjusted R-squared:  0.0504
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 345 weights are ~= 1. The remaining 3668 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0153 0.8640 0.9510 0.9010 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.49e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.215 1          1.102
## LastAuthorFemale 1.211 1          1.101
## Year              1.035 16          1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3152 -0.3487  0.0301  0.3386  2.1036
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9124    0.0511   17.86 < 2e-16 ***
## FirstAuthorFemale1 -0.0485    0.0187   -2.60  0.00943 **
## LastAuthorFemale1 -0.0180    0.0187   -0.96  0.33627
## Year1997         0.0422    0.0712    0.59  0.55304
## Year1998         0.1100    0.0712    1.55  0.12221
## Year1999         0.1150    0.0659    1.74  0.08131 .
## Year2000         0.1942    0.0690    2.81  0.00492 **
## Year2001         0.1641    0.0636    2.58  0.00986 **
## Year2002         0.3016    0.0601    5.02  5.4e-07 ***
## Year2003         0.2925    0.0629    4.65  3.4e-06 ***
## Year2004         0.3459    0.0641    5.40  7.2e-08 ***
## Year2005         0.3246    0.0630    5.15  2.7e-07 ***
```

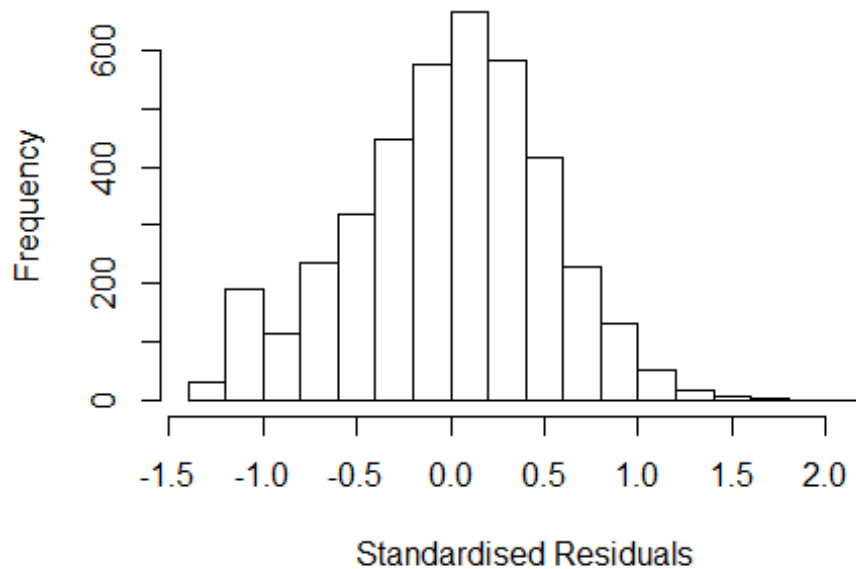


```

## Year2006          0.4028      0.0595      6.77  1.5e-11 ***
## Year2007          0.2478      0.0582      4.25  2.1e-05 ***
## Year2008          0.3234      0.0579      5.59  2.4e-08 ***
## Year2009          0.2289      0.0592      3.87  0.00011 ***
## Year2010          0.1897      0.0573      3.31  0.00094 ***
## Year2011          0.2187      0.0575      3.80  0.00015 ***
## Year2012          0.3045      0.0577      5.28  1.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.0332, Adjusted R-squared:  0.0288
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 323 weights are ~= 1. The remaining 3690 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0453 0.8630 0.9490 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.011
## Year              1.023 16      1.001

```

## Residuals from first author



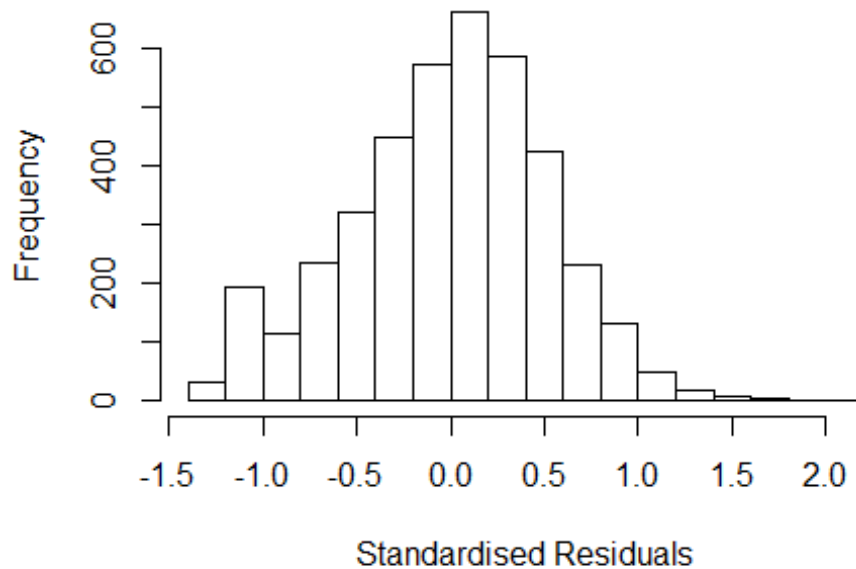
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3119 -0.3479 0.0312 0.3416 2.0988
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9087 0.0509 17.87 < 2e-16 ***
## FirstAuthorFemale1 -0.0573 0.0172 -3.33 0.00087 ***
## Year1997 0.0417 0.0713 0.59 0.55850
## Year1998 0.1117 0.0712 1.57 0.11669
## Year1999 0.1147 0.0660 1.74 0.08248 .
## Year2000 0.1941 0.0691 2.81 0.00500 **
## Year2001 0.1642 0.0636 2.58 0.00985 **
## Year2002 0.3021 0.0602 5.02 5.4e-07 ***
## Year2003 0.2922 0.0629 4.64 3.5e-06 ***
## Year2004 0.3441 0.0641 5.37 8.2e-08 ***
## Year2005 0.3235 0.0630 5.13 3.0e-07 ***
## Year2006 0.4032 0.0595 6.77 1.4e-11 ***
```

```

## Year2007          0.2473      0.0583      4.24  2.3e-05 ***
## Year2008          0.3233      0.0579      5.58  2.5e-08 ***
## Year2009          0.2290      0.0593      3.86  0.00011 ***
## Year2010          0.1895      0.0574      3.30  0.00097 ***
## Year2011          0.2187      0.0576      3.80  0.00015 ***
## Year2012          0.3038      0.0577      5.26  1.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.0329, Adjusted R-squared:  0.0288
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 321 weights are ~= 1. The remaining 3692 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0466 0.8650 0.9490 0.9010 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.019 1      1.009
## Year              1.019 16      1.001

```

## Residuals from last author



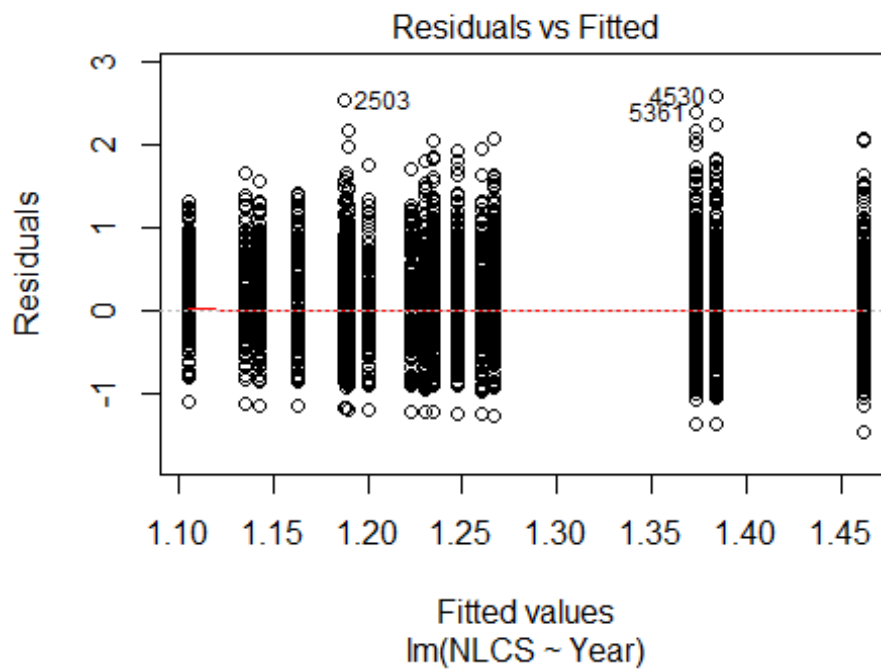
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3012 -0.3443  0.0277  0.3411  2.0936
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9014    0.0512   17.62 < 2e-16 ***
## LastAuthorFemale1 -0.0421    0.0173   -2.43  0.01513 *
## Year1997        0.0458    0.0714    0.64  0.52069
## Year1998        0.1122    0.0713    1.57  0.11586
## Year1999        0.1138    0.0663    1.71  0.08648 .
## Year2000        0.1937    0.0693    2.79  0.00522 **
## Year2001        0.1637    0.0638    2.57  0.01031 *
## Year2002        0.2991    0.0602    4.97  7.1e-07 ***
## Year2003        0.2920    0.0630    4.64  3.6e-06 ***
## Year2004        0.3449    0.0643    5.36  8.7e-08 ***
## Year2005        0.3250    0.0633    5.14  2.9e-07 ***
## Year2006        0.3999    0.0598    6.68  2.6e-11 ***
```

```

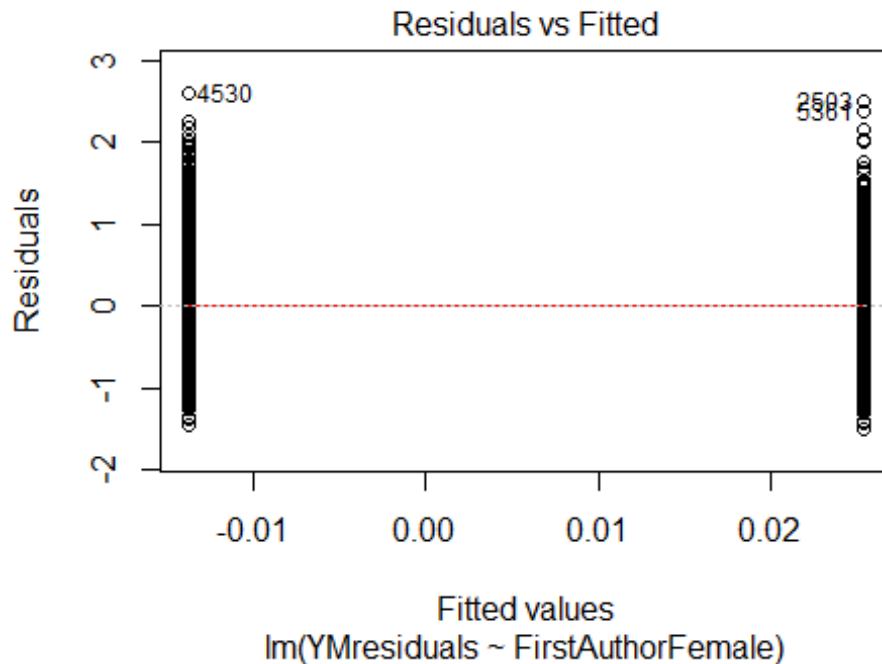
## Year2007          0.2480      0.0585      4.24  2.3e-05 ***
## Year2008          0.3240      0.0582      5.57  2.8e-08 ***
## Year2009          0.2263      0.0594      3.81  0.00014 ***
## Year2010          0.1885      0.0576      3.27  0.00107 **
## Year2011          0.2166      0.0578      3.75  0.00018 ***
## Year2012          0.3011      0.0579      5.20  2.1e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.506
## Multiple R-squared:  0.0316, Adjusted R-squared:  0.0275
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 328 weights are ~= 1. The remaining 3685 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0488 0.8660 0.9490 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.49e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4013"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3300"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 189 202 231 241 266 250 263 201 259 269 370 423 399 433 468
## 2011 2012
## 460 461
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 150 170 181 205 203 187 224 172 216 226 316 353 336 383 412
## 2011 2012

```

```
## 408 410
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 145 162 176 202 198 181 218 167 213 219 311 338 319 364 394
## 2011 2012
## 395 397
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 43, df = 16, p-value = 2e-04
```

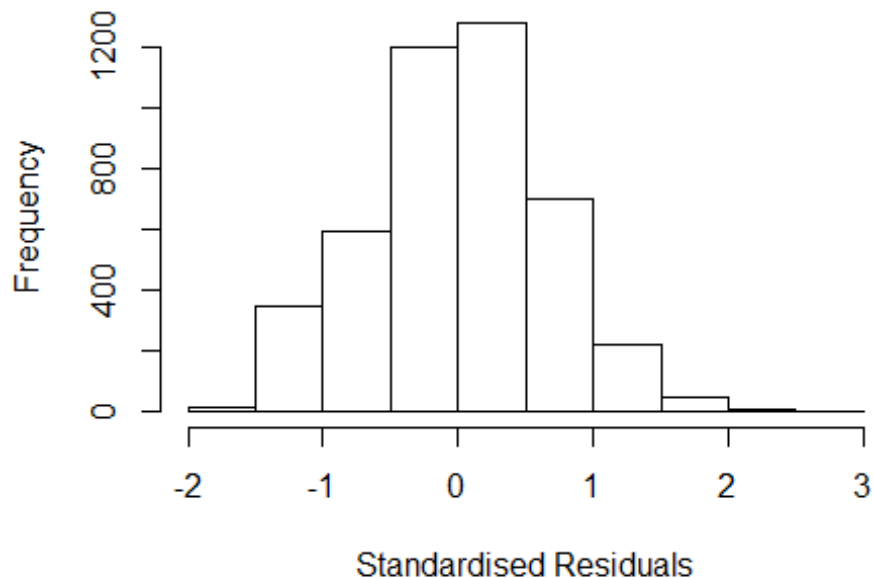


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.8, df = 1, p-value = 0.09
```



```
## [1] "Female first author team size 2018 geometric mean: 1.56044160628882"
## [1] "Male first author team size 2018 geometric mean: 1.44151392007206"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 19000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.55278572184889"
## [1] "Male last author team size 2018 geometric mean: 1.44910089330026"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 18000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.958 1          1.399
## LastAuthorFemale  1.956 1          1.398
## UniqueAuthors    1.064 4          1.008
## Year              1.097 16         1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 4530 77957319670 3.965 2010      3300      1      2.698
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6501 -0.4241  0.0155  0.4336  2.6978
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.14847    0.05862   19.59 < 2e-16 ***
## FirstAuthorFemale1 0.02971    0.02905    1.02  0.30646
## LastAuthorFemale1 0.00774    0.02929    0.26  0.79153
## UniqueAuthors2    0.17873    0.02604    6.86  7.7e-12 ***
## UniqueAuthors3    0.22936    0.03433    6.68  2.7e-11 ***
## UniqueAuthors4    0.25529    0.04694    5.44  5.7e-08 ***
## UniqueAuthors5    0.22262    0.07867    2.83  0.00468 **
## Year1997         -0.05494    0.07942   -0.69  0.48912
## Year1998         -0.06001    0.07745   -0.77  0.43847
## Year1999         -0.08457    0.07059   -1.20  0.23096
```

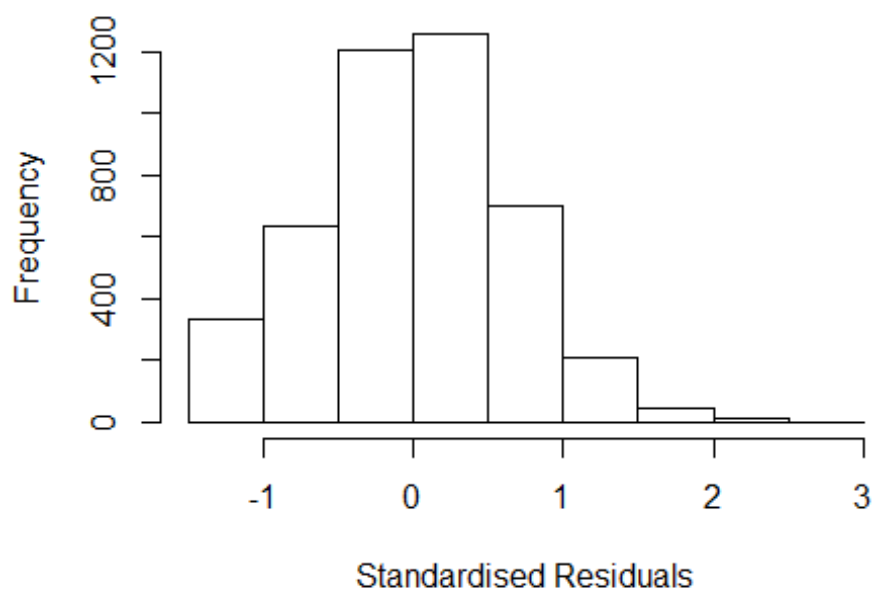


```

## Year2000      -0.10753    0.07513   -1.43  0.15241
## Year2001      -0.11095    0.07711   -1.44  0.15027
## Year2002      -0.03893    0.07228   -0.54  0.59018
## Year2003      -0.02801    0.07375   -0.38  0.70413
## Year2004       0.05074    0.07249    0.70  0.48399
## Year2005      -0.04567    0.07058   -0.65  0.51765
## Year2006       0.01305    0.06774    0.19  0.84725
## Year2007       0.02083    0.06852    0.30  0.76111
## Year2008       0.01600    0.06813    0.23  0.81439
## Year2009       0.04580    0.06790    0.67  0.50004
## Year2010       0.11870    0.06740    1.76  0.07830 .
## Year2011       0.13764    0.07089    1.94  0.05224 .
## Year2012       0.23479    0.06788    3.46  0.00055 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.642
## Multiple R-squared:  0.0431, Adjusted R-squared:  0.0383
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 377 weights are ~= 1. The remaining 4022 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0382 0.8650 0.9500 0.9040 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.27e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.006 1 1.416
## LastAuthorFemale 1.999 1 1.414
## Year 1.042 16 1.001

```

## Residuals from first and last author



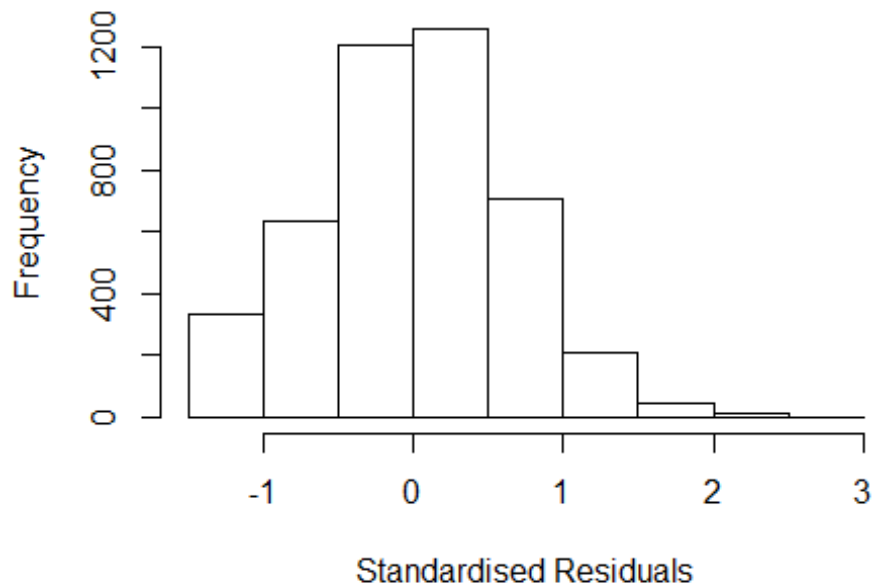
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2503 14644388070 3.717 2005      3300      1      2.51
## 4530 77957319670 3.965 2010      3300      1      2.63
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4952 -0.4305  0.0121  0.4408  2.6297
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.208149   0.059521  20.30 < 2e-16 ***
## FirstAuthorFemale1  0.047276   0.029760   1.59  0.11223
## LastAuthorFemale1 -0.000626   0.030002  -0.02  0.98336
## Year1997        -0.056328   0.080412  -0.70  0.48366
## Year1998        -0.060581   0.079229  -0.76  0.44453
## Year1999        -0.084622   0.072115  -1.17  0.24069
## Year2000        -0.098193   0.077254  -1.27  0.20378
## Year2001        -0.114935   0.078057  -1.47  0.14097
## Year2002        -0.044574   0.073502  -0.61  0.54426
## Year2003        -0.028031   0.074847  -0.37  0.70804
## Year2004         0.046058   0.073726   0.62  0.53218
```

```

## Year2005      -0.048152   0.072115   -0.67   0.50435
## Year2006      0.010499   0.069160    0.15   0.87935
## Year2007      0.023191   0.069809    0.33   0.73975
## Year2008      0.011478   0.069817    0.16   0.86942
## Year2009      0.045603   0.069514    0.66   0.51184
## Year2010      0.127198   0.068858    1.85   0.06478 .
## Year2011      0.145312   0.072314    2.01   0.04455 *
## Year2012      0.239745   0.069148    3.47   0.00053 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0234, Adjusted R-squared:  0.0194
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 397 weights are ~= 1. The remaining 4002 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0648 0.8630 0.9500 0.9040 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.022 1      1.011
## Year      1.022 16      1.001

```

## Residuals from first author



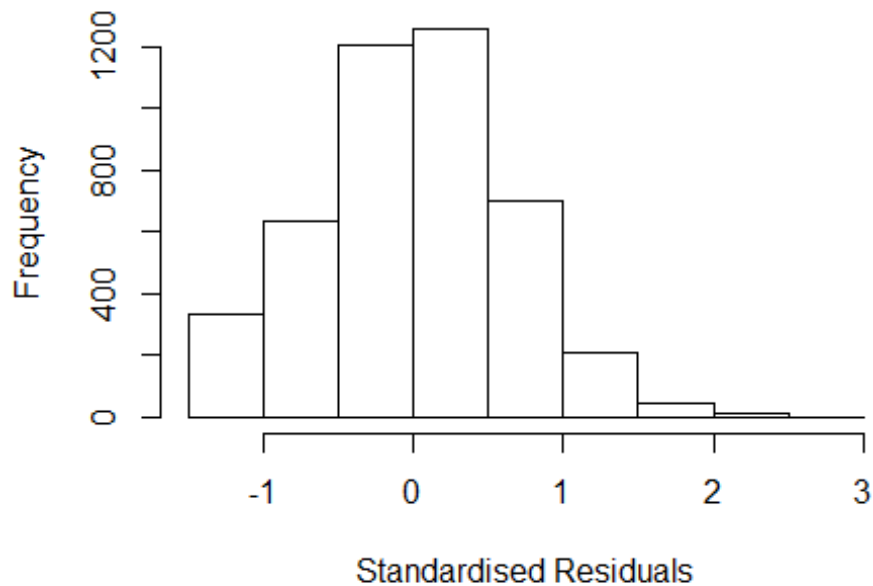
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2503 14644388070 3.717 2005      3300      1      2.51
## 4530 77957319670 3.965 2010      3300      1      2.63
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4947 -0.4305  0.0122  0.4409  2.6297
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2081     0.0594   20.32 < 2e-16 ***
## FirstAuthorFemale1  0.0468     0.0213    2.20  0.02758 *
## Year1997         -0.0563     0.0804   -0.70  0.48380
## Year1998         -0.0606     0.0792   -0.76  0.44445
## Year1999         -0.0846     0.0721   -1.17  0.24069
## Year2000         -0.0982     0.0773   -1.27  0.20386
## Year2001         -0.1149     0.0780   -1.47  0.14095
## Year2002         -0.0446     0.0735   -0.61  0.54436
## Year2003         -0.0280     0.0748   -0.37  0.70789
## Year2004          0.0461     0.0737    0.62  0.53209
## Year2005         -0.0481     0.0721   -0.67  0.50435
```

```

## Year2006          0.0105      0.0692      0.15  0.87944
## Year2007          0.0232      0.0698      0.33  0.73989
## Year2008          0.0115      0.0698      0.16  0.86927
## Year2009          0.0456      0.0695      0.66  0.51194
## Year2010          0.1272      0.0689      1.85  0.06485 .
## Year2011          0.1453      0.0723      2.01  0.04457 *
## Year2012          0.2397      0.0691      3.47  0.00053 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0234, Adjusted R-squared:  0.0196
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 397 weights are ~= 1. The remaining 4002 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0646 0.8620 0.9500 0.9040 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.019 1      1.009
## Year      1.019 16      1.001

```

## Residuals from last author



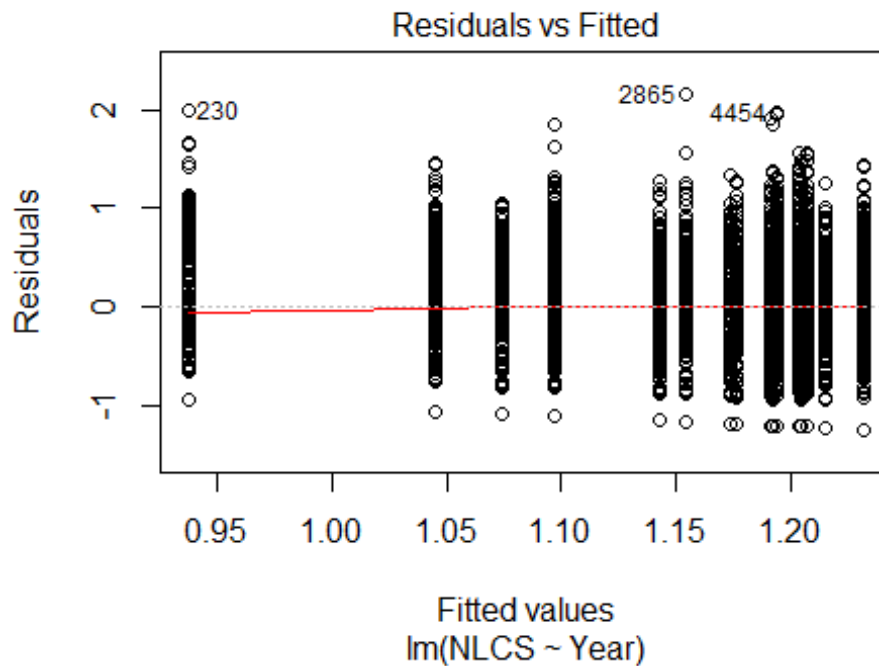
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2503 14644388070 3.717 2005      3300      1      2.51
## 4530 77957319670 3.965 2010      3300      1      2.63
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4861 -0.4349  0.0134  0.4396  2.6225
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2108     0.0597  20.27 < 2e-16 ***
## LastAuthorFemale1  0.0333     0.0214   1.56  0.11970
## Year1997        -0.0539     0.0806  -0.67  0.50391
## Year1998        -0.0603     0.0794  -0.76  0.44789
## Year1999        -0.0834     0.0723  -1.15  0.24893
## Year2000        -0.0977     0.0775  -1.26  0.20748
## Year2001        -0.1103     0.0781  -1.41  0.15809
## Year2002        -0.0406     0.0736  -0.55  0.58124
## Year2003        -0.0269     0.0751  -0.36  0.71986
## Year2004         0.0485     0.0739   0.66  0.51155
## Year2005        -0.0444     0.0722  -0.61  0.53867
```

```

## Year2006          0.0127      0.0694      0.18  0.85476
## Year2007          0.0248      0.0700      0.35  0.72294
## Year2008          0.0142      0.0701      0.20  0.83921
## Year2009          0.0493      0.0697      0.71  0.47930
## Year2010          0.1317      0.0689      1.91  0.05616 .
## Year2011          0.1485      0.0724      2.05  0.04036 *
## Year2012          0.2420      0.0694      3.49  0.00049 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.65
## Multiple R-squared:  0.0229, Adjusted R-squared:  0.0191
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 389 weights are ~= 1. The remaining 4010 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0672 0.8620 0.9500 0.9050 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.27e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4399"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3301"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 224 204 234 224 298 278 263 258 244 288 302 451 389 415 468
## 2011 2012
## 454 494
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 196 165 186 191 238 216 211 207 203 245 255 379 331 360 406

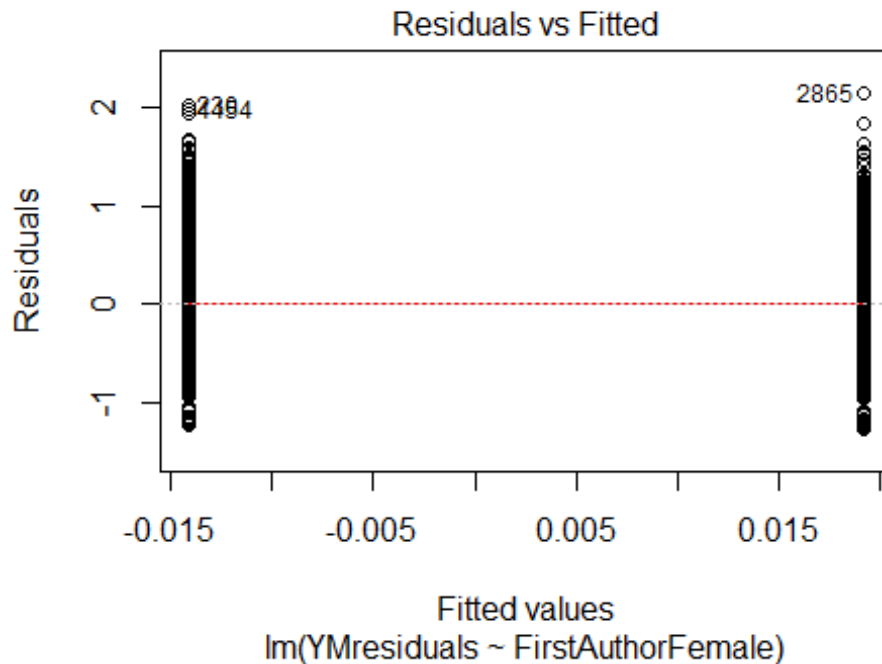
```

```
## 2011 2012
## 391 430
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 192 162 180 187 226 208 206 200 191 239 240 358 316 342 389
## 2011 2012
## 375 415
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 46, df = 16, p-value = 1e-04
```



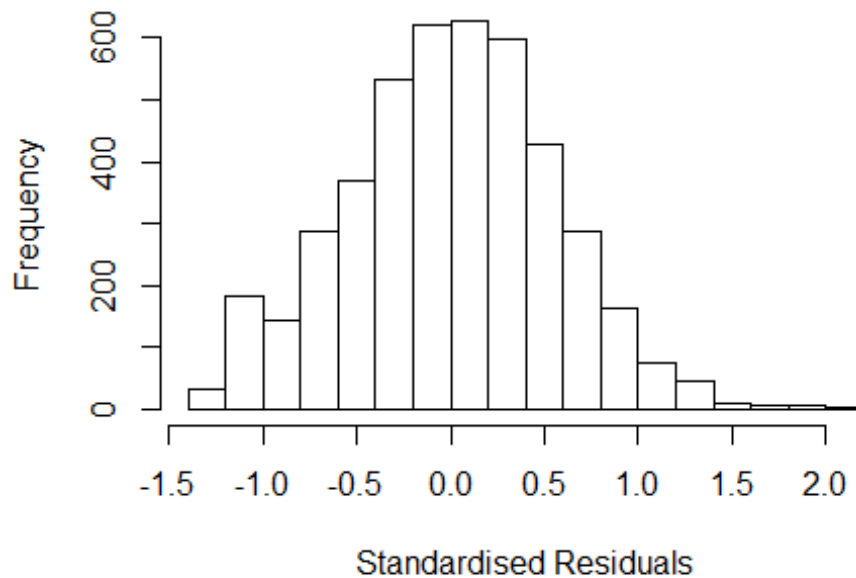
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.57, df = 1, p-value = 0.5
```





```
## [1] "Female first author team size 2018 geometric mean: 1.95963501723337"
## [1] "Male first author team size 2018 geometric mean: 1.57390362206099"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 29000, p-value = 7e-05
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.90512003067845"
## [1] "Male last author team size 2018 geometric mean: 1.63103457960435"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 28000, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.704 1      1.305
## LastAuthorFemale  1.688 1      1.299
## UniqueAuthors     1.078 4      1.009
## Year               1.078 16     1.002
```

## Residuals from first and last author and team size



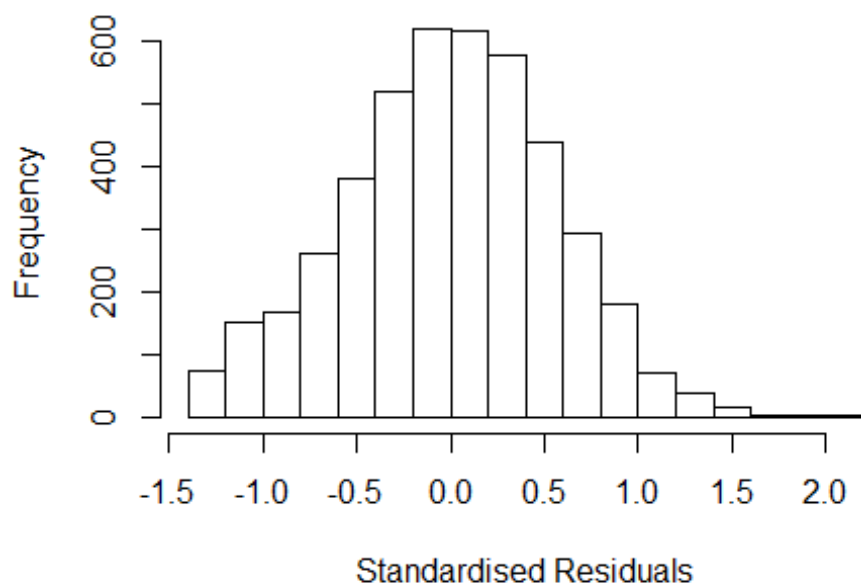
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.348 -0.364  0.010  0.377  2.195
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8350    0.0500   16.71 < 2e-16 ***
## FirstAuthorFemale1 0.0179    0.0226    0.79  0.42884
## LastAuthorFemale1 0.0149    0.0225    0.66  0.50768
## UniqueAuthors2    0.1193    0.0204    5.85  5.1e-09 ***
## UniqueAuthors3    0.1794    0.0288    6.24  4.8e-10 ***
## UniqueAuthors4    0.2142    0.0374    5.73  1.1e-08 ***
## UniqueAuthors5    0.2394    0.0471    5.09  3.8e-07 ***
## Year1997          0.1510    0.0656    2.30  0.02147 *
## Year1998          0.1317    0.0666    1.98  0.04792 *
## Year1999          0.2478    0.0655    3.78  0.00016 ***
```

```

## Year2000          0.1744      0.0620      2.81  0.00491 **
## Year2001          0.2776      0.0635      4.37  1.2e-05 ***
## Year2002          0.3020      0.0634      4.76  2.0e-06 ***
## Year2003          0.3150      0.0614      5.13  3.1e-07 ***
## Year2004          0.2708      0.0620      4.37  1.3e-05 ***
## Year2005          0.2640      0.0587      4.50  7.1e-06 ***
## Year2006          0.2428      0.0599      4.05  5.1e-05 ***
## Year2007          0.1976      0.0593      3.33  0.00086 ***
## Year2008          0.2899      0.0574      5.05  4.7e-07 ***
## Year2009          0.2689      0.0581      4.62  3.9e-06 ***
## Year2010          0.3142      0.0561      5.60  2.2e-08 ***
## Year2011          0.2938      0.0597      4.92  8.9e-07 ***
## Year2012          0.3004      0.0586      5.13  3.0e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.0394, Adjusted R-squared:  0.0346
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 380 weights are ~= 1. The remaining 4046 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0755 0.8650 0.9510 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.26e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.698 1          1.303
## LastAuthorFemale 1.700 1          1.304
## Year          1.026 16          1.001

```

## Residuals from first and last author



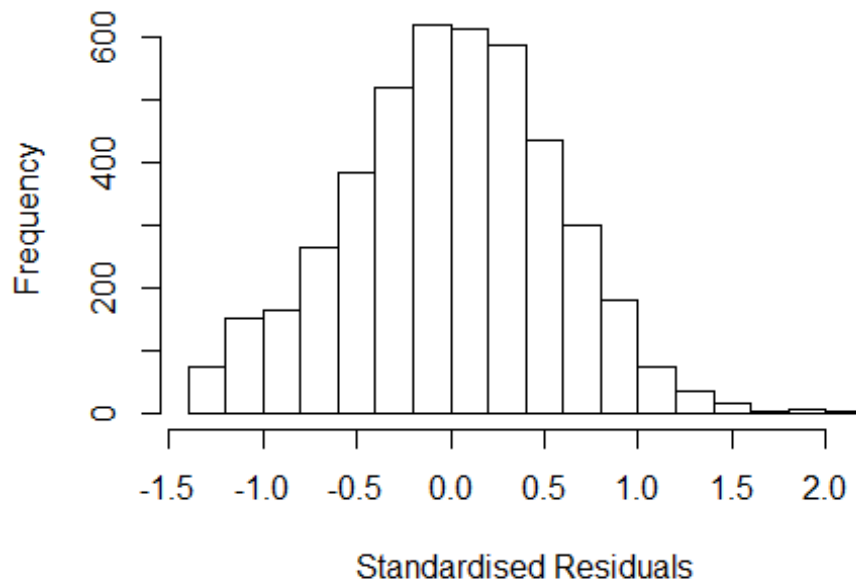
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2593 -0.3737  0.0125  0.3792  2.1232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8780    0.0499   17.58 < 2e-16 ***
## FirstAuthorFemale1 0.0364    0.0228    1.60  0.11018
## LastAuthorFemale1 0.0137    0.0228    0.60  0.54933
## Year1997        0.1533    0.0652    2.35  0.01885 *
## Year1998        0.1308    0.0672    1.95  0.05163 .
## Year1999        0.2571    0.0659    3.90  9.6e-05 ***
## Year2000        0.1887    0.0621    3.04  0.00237 **
## Year2001        0.2889    0.0637    4.53  6.0e-06 ***
## Year2002        0.3141    0.0630    4.98  6.5e-07 ***
## Year2003        0.3313    0.0615    5.39  7.6e-08 ***
## Year2004        0.2838    0.0625    4.54  5.7e-06 ***
## Year2005        0.2782    0.0598    4.66  3.3e-06 ***
```

```

## Year2006          0.2548      0.0602      4.23  2.4e-05 ***
## Year2007          0.2073      0.0600      3.46  0.00055 ***
## Year2008          0.3017      0.0577      5.23  1.8e-07 ***
## Year2009          0.2794      0.0584      4.79  1.8e-06 ***
## Year2010          0.3265      0.0562      5.81  6.6e-09 ***
## Year2011          0.3082      0.0598      5.16  2.6e-07 ***
## Year2012          0.3167      0.0589      5.37  8.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0206, Adjusted R-squared:  0.0166
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 363 weights are ~= 1. The remaining 4063 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.111  0.867   0.951   0.906   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1          1.006
## Year              1.012 16          1.000

```

## Residuals from first author



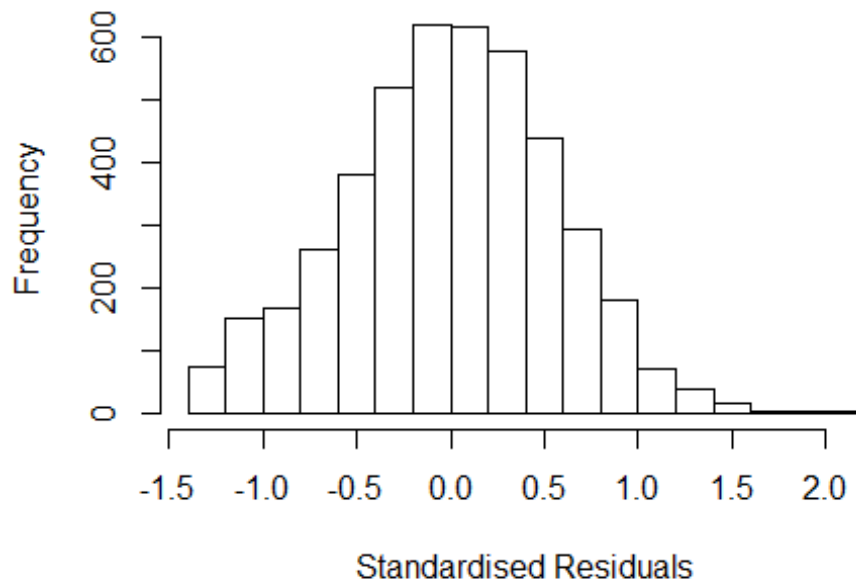
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2566 -0.3762  0.0112  0.3788  2.1252
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8797    0.0497   17.69 < 2e-16 ***
## FirstAuthorFemale1 0.0455    0.0176    2.58 0.00978 **
## Year1997        0.1530    0.0652    2.35 0.01892 *
## Year1998        0.1307    0.0672    1.95 0.05170 .
## Year1999        0.2575    0.0658    3.91 9.3e-05 ***
## Year2000        0.1890    0.0620    3.05 0.00232 **
## Year2001        0.2890    0.0637    4.53 5.9e-06 ***
## Year2002        0.3144    0.0630    4.99 6.3e-07 ***
## Year2003        0.3314    0.0615    5.39 7.4e-08 ***
## Year2004        0.2839    0.0624    4.55 5.5e-06 ***
## Year2005        0.2785    0.0597    4.66 3.2e-06 ***
## Year2006        0.2556    0.0602    4.25 2.2e-05 ***
```

```

## Year2007          0.2072      0.0599      3.46  0.00055 ***
## Year2008          0.3017      0.0577      5.23  1.8e-07 ***
## Year2009          0.2791      0.0583      4.78  1.8e-06 ***
## Year2010          0.3266      0.0561      5.82  6.4e-09 ***
## Year2011          0.3077      0.0597      5.15  2.7e-07 ***
## Year2012          0.3171      0.0589      5.38  7.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0205, Adjusted R-squared:  0.0168
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 368 weights are ~= 1. The remaining 4058 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.110  0.868  0.951  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.013 1      1.007
## Year              1.013 16      1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2528 -0.3776 0.0116 0.3770 2.1314
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8815 0.0498 17.68 < 2e-16 ***
## LastAuthorFemale1 0.0379 0.0176 2.15 0.03163 *
## Year1997 0.1546 0.0653 2.37 0.01789 *
## Year1998 0.1318 0.0671 1.96 0.04974 *
## Year1999 0.2575 0.0657 3.92 9.0e-05 ***
## Year2000 0.1910 0.0620 3.08 0.00209 **
## Year2001 0.2890 0.0637 4.54 5.8e-06 ***
## Year2002 0.3146 0.0629 5.00 6.0e-07 ***
## Year2003 0.3334 0.0615 5.42 6.2e-08 ***
## Year2004 0.2861 0.0625 4.58 4.8e-06 ***
## Year2005 0.2797 0.0598 4.68 2.9e-06 ***
## Year2006 0.2552 0.0601 4.24 2.2e-05 ***
```

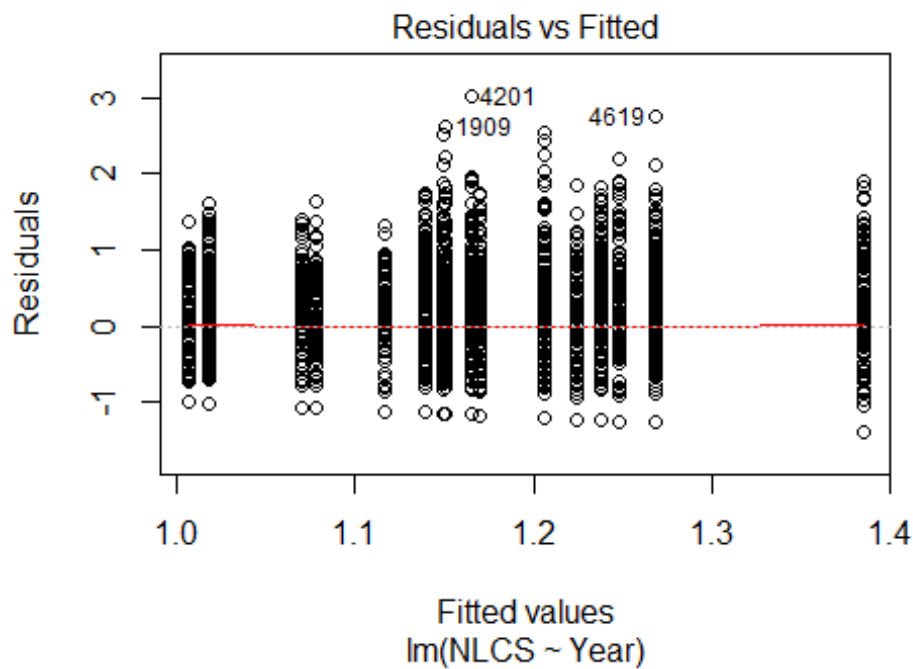


```

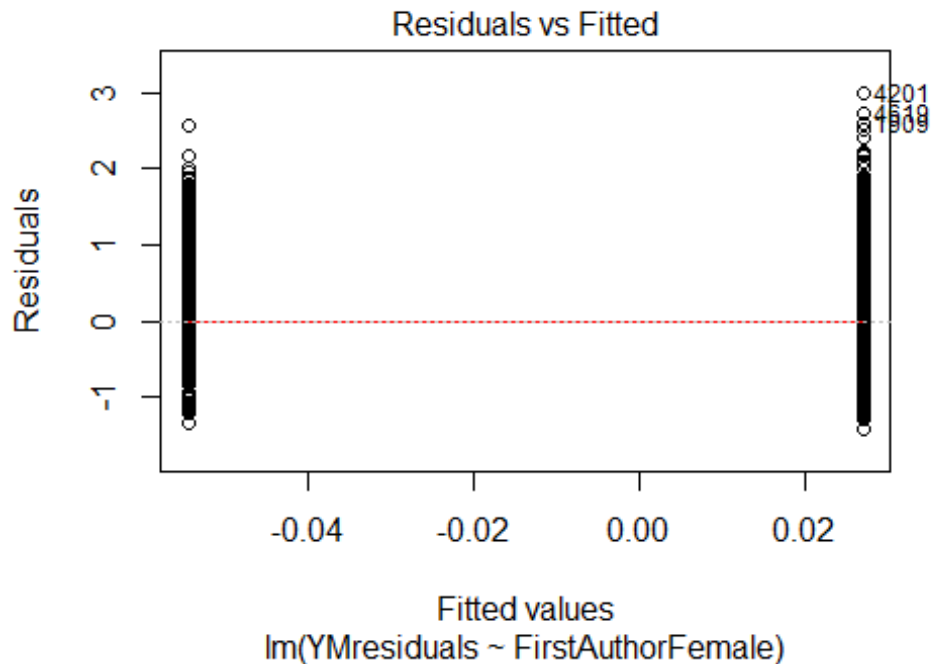
## Year2007          0.2097      0.0599      3.50  0.00046 ***
## Year2008          0.3043      0.0577      5.28  1.4e-07 ***
## Year2009          0.2824      0.0583      4.84  1.3e-06 ***
## Year2010          0.3287      0.0562      5.85  5.2e-09 ***
## Year2011          0.3116      0.0597      5.22  1.9e-07 ***
## Year2012          0.3183      0.0589      5.40  6.9e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.555
## Multiple R-squared:  0.0201, Adjusted R-squared:  0.0163
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 370 weights are ~= 1. The remaining 4056 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.108  0.868  0.951  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4426"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3302"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 172 191 188 168 210 182 229 216 240 215 248 260 277 232 255
## 2011 2012
## 292 306
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 115 118 135 116 138 113 173 164 195 161 198 199 222 182 188
## 2011 2012

```

```
## 232 252
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 114 117 132 113 133 110 171 157 184 153 186 193 214 169 181
## 2011 2012
## 223 246
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 74, df = 16, p-value = 2e-09
```

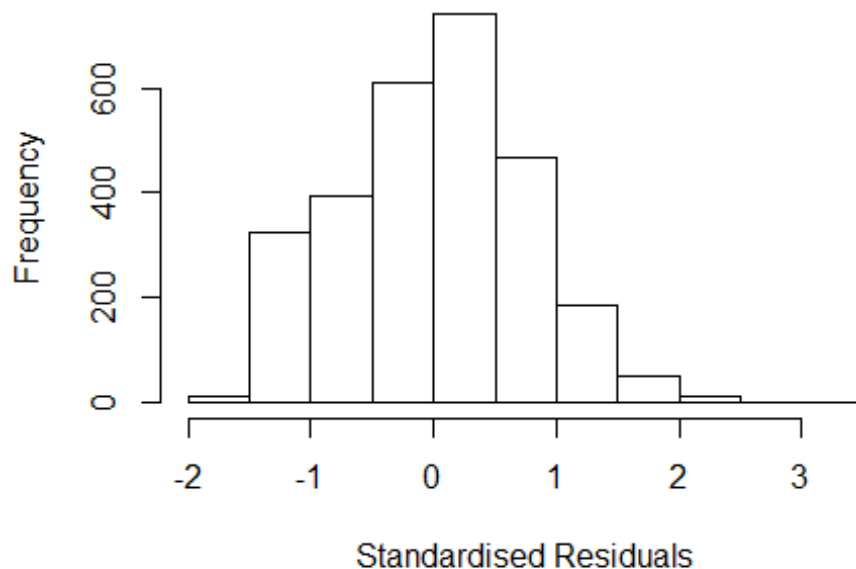


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.4, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.59903501055078"
## [1] "Male first author team size 2018 geometric mean: 1.38900770961058"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7200, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.47843941437996"
## [1] "Male last author team size 2018 geometric mean: 1.45199830458861"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.404 1      1.551
## LastAuthorFemale 2.462 1      1.569
## UniqueAuthors    1.181 4      1.021
## Year             1.178 16     1.005
```

## Residuals from first and last author and team size



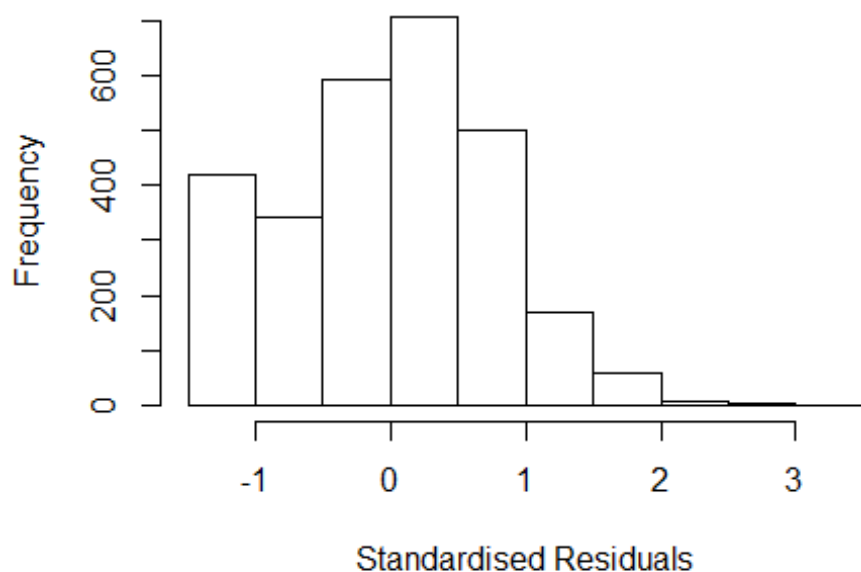
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2765 33947152939 3.747 2007      1200      3      2.644
## 4201 79951880145 4.181 2011      1200      2      3.140
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7078 -0.5408  0.0325  0.5095  3.1400
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.19101    0.06792   17.54 < 2e-16 ***
## FirstAuthorFemale1 -0.00788    0.04913   -0.16  0.87258
## LastAuthorFemale1 -0.10662    0.05088   -2.10  0.03621 *
## UniqueAuthors2     0.34702    0.04506    7.70 1.9e-14 ***
## UniqueAuthors3     0.47806    0.05484    8.72 < 2e-16 ***
## UniqueAuthors4     0.56999    0.06991    8.15 5.3e-16 ***
## UniqueAuthors5     0.48410    0.07099    6.82 1.1e-11 ***
## Year1997        -0.16780    0.09324   -1.80  0.07201 .
## Year1998        -0.22821    0.08736   -2.61  0.00904 **
```

```

## Year1999      -0.15186      0.09208      -1.65      0.09921 .
## Year2000      -0.14843      0.08805      -1.69      0.09196 .
## Year2001      -0.24610      0.09156      -2.69      0.00724 **
## Year2002      -0.05169      0.09842      -0.53      0.59953
## Year2003      -0.12878      0.09075      -1.42      0.15597
## Year2004      -0.10884      0.08853      -1.23      0.21905
## Year2005       0.13342      0.09232       1.45      0.14853
## Year2006      -0.04131      0.09085      -0.45      0.64933
## Year2007      -0.08781      0.09447      -0.93      0.35271
## Year2008      -0.14426      0.08548      -1.69      0.09159 .
## Year2009      -0.16334      0.08813      -1.85      0.06393 .
## Year2010      -0.29345      0.08406      -3.49      0.00049 ***
## Year2011      -0.15000      0.08696      -1.72      0.08467 .
## Year2012      -0.04534      0.08659      -0.52      0.60063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.77
## Multiple R-squared:  0.0784, Adjusted R-squared:  0.071
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 263 weights are ~= 1. The remaining 2533 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0591 0.8630 0.9480 0.9140 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.58e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.641 1 1.625
## LastAuthorFemale 2.679 1 1.637
## Year 1.051 16 1.002

```

## Residuals from first and last author



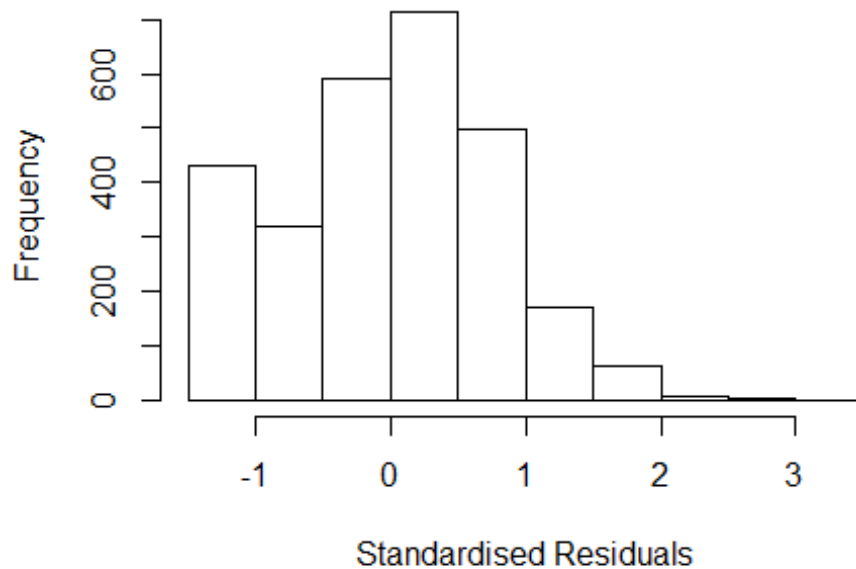
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1909  1942506038 3.776 2004      1200      2      2.805
## 2765  33947152939 3.747 2007      1200      3      2.540
## 2810  34047099783 3.640 2007      1204      2      2.632
## 4201  79951880145 4.181 2011      1200      2      3.014
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4321 -0.5590  0.0469  0.5385  3.0142
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.259158   0.065062   19.35  < 2e-16 ***
## FirstAuthorFemale1  0.066055   0.053228    1.24  0.21472
## LastAuthorFemale1 -0.199327   0.054825   -3.64  0.00028 ***
## Year1997        -0.130975   0.090442   -1.45  0.14768
## Year1998        -0.214653   0.084883   -2.53  0.01150 *
## Year1999        -0.107995   0.088833   -1.22  0.22420
## Year2000        -0.135158   0.085497   -1.58  0.11402
## Year2001        -0.203224   0.091252   -2.23  0.02602 *
## Year2002        -0.000391   0.098432    0.00  0.99683
```

```

## Year2003      -0.096530    0.089010    -1.08    0.27824
## Year2004      -0.088716    0.086784    -1.02    0.30675
## Year2005       0.172974    0.091690     1.89    0.05933 .
## Year2006       0.015447    0.090951     0.17    0.86515
## Year2007      -0.052123    0.094141    -0.55    0.57985
## Year2008      -0.112969    0.086039    -1.31    0.18929
## Year2009      -0.082648    0.088076    -0.94    0.34813
## Year2010      -0.213492    0.082921    -2.57    0.01009 *
## Year2011      -0.092362    0.085830    -1.08    0.28197
## Year2012       0.023382    0.086300     0.27    0.78646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.79
## Multiple R-squared:  0.0214, Adjusted R-squared:  0.0151
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 230 weights are ~= 1. The remaining 2566 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.114  0.856  0.950   0.914   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.012 1          1.006
## Year              1.012 16          1.000

```

## Residuals from first author



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1909  1942506038 3.776 2004      1200      2      2.805
## 2765  33947152939 3.747 2007      1200      3      2.540
## 2810  34047099783 3.640 2007      1204      2      2.632
## 4201  79951880145 4.181 2011      1200      2      3.014
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4111 -0.5566  0.0503  0.5331  3.0243
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2376    0.0654   18.92  <2e-16 ***
## FirstAuthorFemale1 -0.0852    0.0334   -2.55   0.011 *
## Year1997         -0.1292    0.0907   -1.42   0.154
## Year1998         -0.1984    0.0854   -2.32   0.020 *
## Year1999         -0.0920    0.0892   -1.03   0.303
## Year2000         -0.1269    0.0858   -1.48   0.139
## Year2001         -0.1877    0.0917   -2.05   0.041 *
## Year2002          0.0108    0.0996    0.11   0.914
## Year2003         -0.0908    0.0893   -1.02   0.309
```

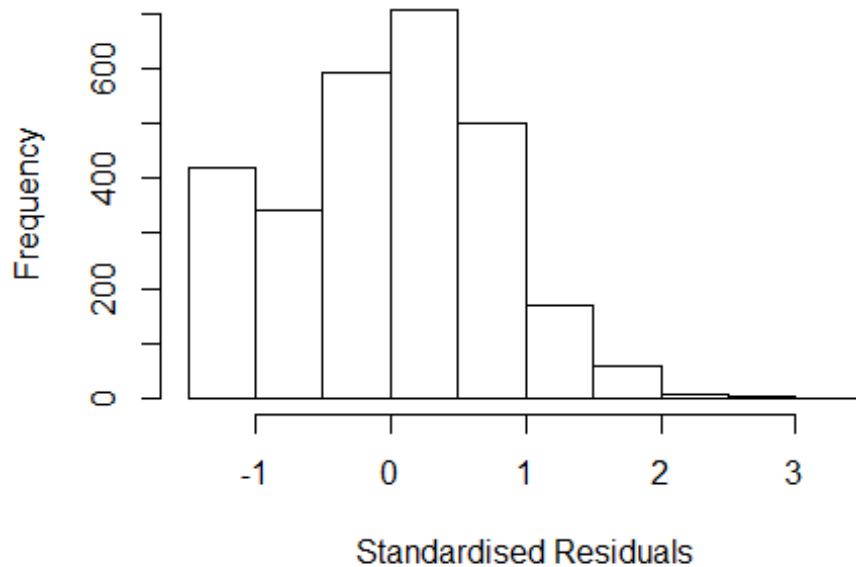


```

## Year2004          -0.0750      0.0872   -0.86    0.390
## Year2005           0.1735      0.0926    1.87    0.061 .
## Year2006           0.0279      0.0918    0.30    0.761
## Year2007          -0.0421      0.0943   -0.45    0.655
## Year2008          -0.0974      0.0869   -1.12    0.263
## Year2009          -0.0653      0.0887   -0.74    0.462
## Year2010          -0.1933      0.0832   -2.32    0.020 *
## Year2011          -0.0809      0.0863   -0.94    0.349
## Year2012           0.0288      0.0863    0.33    0.738
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.791
## Multiple R-squared:  0.0159, Adjusted R-squared:  0.00986
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 231 weights are ~= 1. The remaining 2565 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.111  0.850  0.950   0.914   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.58e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1          1.012
## Year            1.024 16          1.001

```

## Residuals from last author



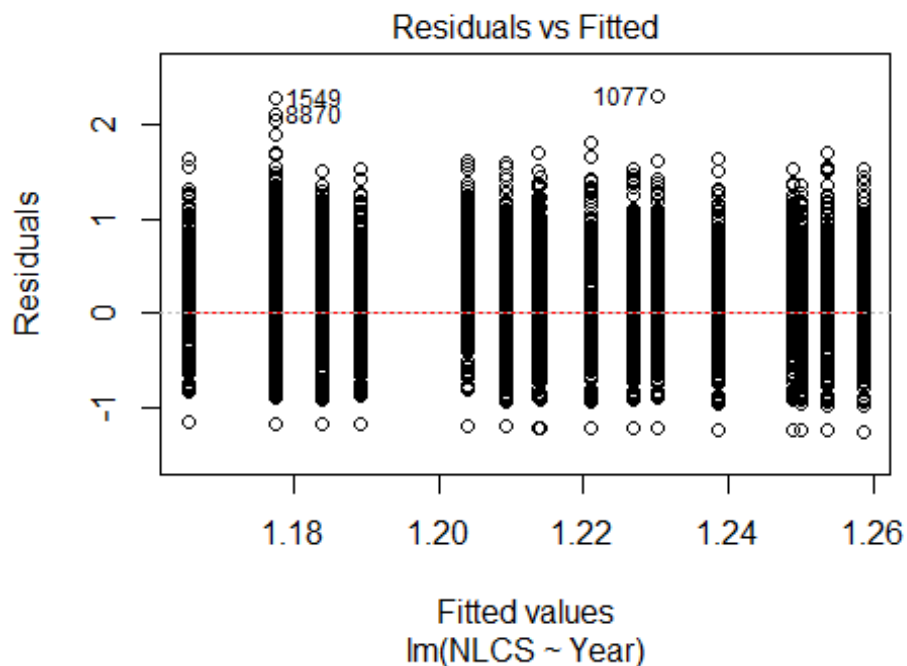
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1909  1942506038 3.776 2004     1200      2      2.805
## 2765  33947152939 3.747 2007     1200      3      2.540
## 2810  34047099783 3.640 2007     1204      2      2.632
## 4201  79951880145 4.181 2011     1200      2      3.014
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4345 -0.5579  0.0454  0.5342  3.0067
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.26058    0.06505   19.38  < 2e-16 ***
## LastAuthorFemale1 -0.14686    0.03379   -4.35  1.4e-05 ***
## Year1997        -0.12868    0.09027   -1.43   0.154
## Year1998        -0.20910    0.08458   -2.47   0.013 *
## Year1999        -0.10311    0.08856   -1.16   0.244
## Year2000        -0.13132    0.08535   -1.54   0.124
## Year2001        -0.19934    0.09112   -2.19   0.029 *
## Year2002         0.00211    0.09859    0.02   0.983
## Year2003        -0.09373    0.08891   -1.05   0.292
```

```

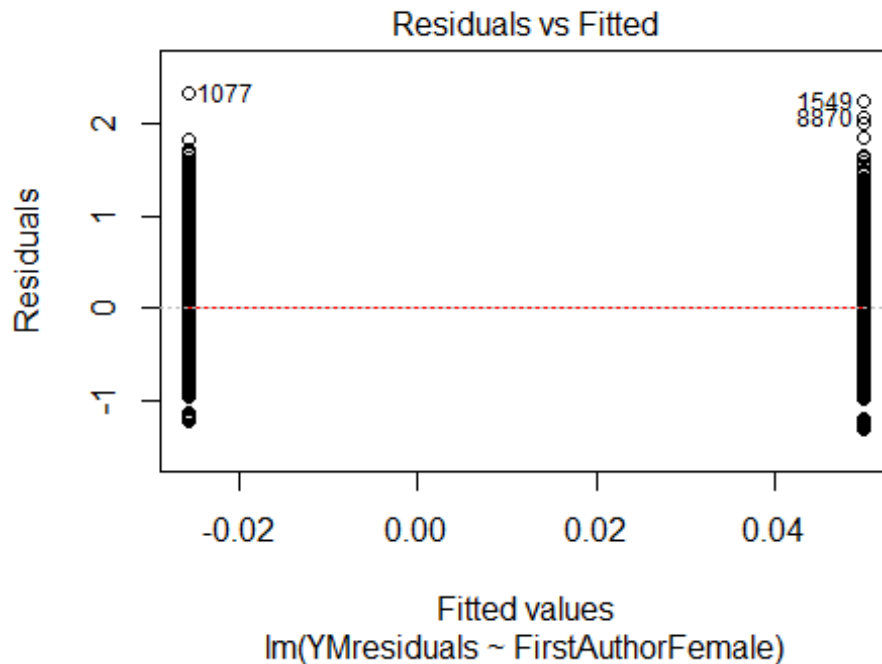
## Year2004          -0.08403      0.08666      -0.97      0.332
## Year2005           0.17394      0.09177       1.90      0.058 .
## Year2006           0.01977      0.09088       0.22      0.828
## Year2007          -0.04830      0.09411      -0.51      0.608
## Year2008          -0.10654      0.08593      -1.24      0.215
## Year2009          -0.07643      0.08810      -0.87      0.386
## Year2010          -0.20600      0.08259      -2.49      0.013 *
## Year2011          -0.08626      0.08570      -1.01      0.314
## Year2012           0.02736      0.08604       0.32      0.751
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.79
## Multiple R-squared:  0.0207, Adjusted R-squared:  0.0147
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 229 weights are ~= 1. The remaining 2567 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.116  0.857  0.950   0.914   0.984   0.999
## Algorithmic parameters:
##           tuning.chi                bb           tuning.psi           refine.tol
##           1.55e+00                5.00e-01           4.69e+00           1.00e-07
##           rel.tol                solve.tol           eps.outlier           eps.x
##           1.00e-07                1.00e-07           3.58e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01                5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi                subsampling                cov
##           "bisquare"                "nonsingular"                ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2796"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3303"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 359 427 400 450 508 466 498 521 527 510 525 487 434 501 544
## 2011 2012
## 677 585
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 301 339 330 341 333 238 419 429 440 441 435 395 373 426 454
## 2011 2012
## 551 480
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 295 330 319 331 321 228 406 409 422 430 418 377 363 409 432
## 2011 2012
## 529 462
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 57, df = 16, p-value = 2e-06
```

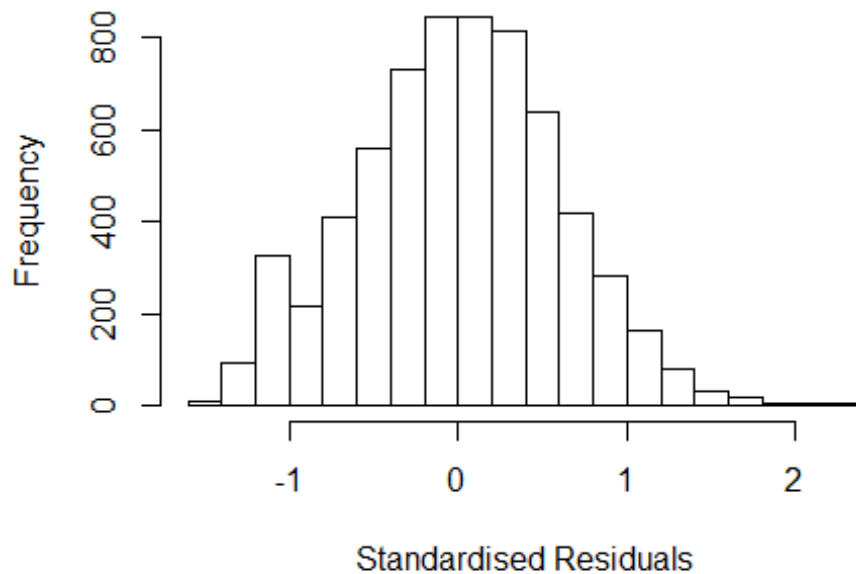


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 0.001
```



```
## [1] "Female first author team size 2018 geometric mean: 1.43562910734814"
## [1] "Male first author team size 2018 geometric mean: 1.26530853739342"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26000, p-value = 0.009
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.34227243127569"
## [1] "Male last author team size 2018 geometric mean: 1.32777658501306"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23000, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.968 1          1.403
## LastAuthorFemale  1.960 1          1.400
## UniqueAuthors    1.058 4          1.007
## Year              1.074 16         1.002
```

## Residuals from first and last author and team size



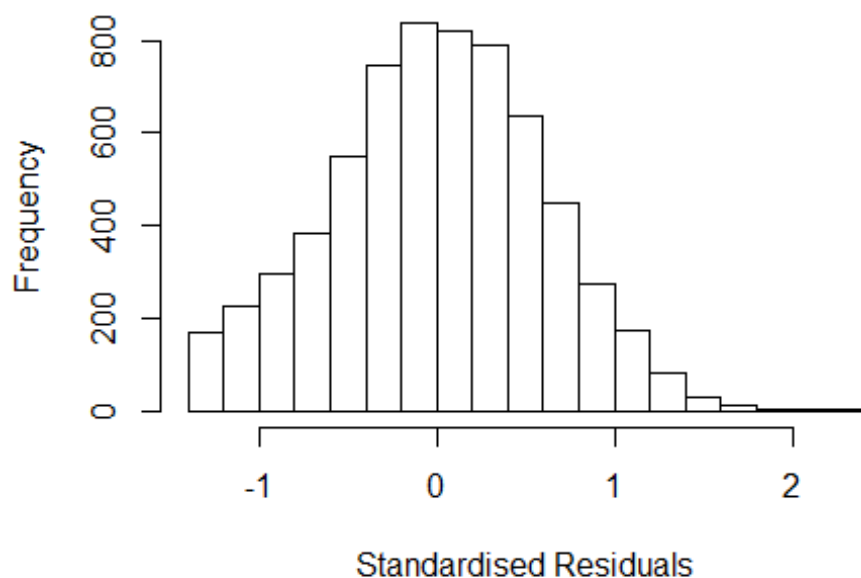
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.46873 -0.39697 0.00826 0.40382 2.37551
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.09122 0.04067 26.83 < 2e-16 ***
## FirstAuthorFemale1 0.06010 0.02212 2.72 0.0066 **
## LastAuthorFemale1 0.02485 0.02213 1.12 0.2616
## UniqueAuthors2 0.16002 0.01916 8.35 < 2e-16 ***
## UniqueAuthors3 0.26327 0.03077 8.56 < 2e-16 ***
## UniqueAuthors4 0.22225 0.05341 4.16 3.2e-05 ***
## UniqueAuthors5 0.31835 0.06651 4.79 1.7e-06 ***
## Year1997 0.05501 0.05618 0.98 0.3276
## Year1998 0.06928 0.05477 1.26 0.2060
## Year1999 -0.02198 0.05610 -0.39 0.6952
```

```

## Year2000      0.07428    0.05382    1.38    0.1675
## Year2001      0.03156    0.05528    0.57    0.5681
## Year2002      0.03555    0.05216    0.68    0.4956
## Year2003      0.02527    0.05005    0.50    0.6137
## Year2004      0.04475    0.05006    0.89    0.3714
## Year2005      0.05415    0.04761    1.14    0.2555
## Year2006      0.08663    0.04786    1.81    0.0703 .
## Year2007      0.05773    0.05044    1.14    0.2525
## Year2008      0.01884    0.05073    0.37    0.7104
## Year2009      0.02158    0.04990    0.43    0.6655
## Year2010      0.06679    0.04971    1.34    0.1791
## Year2011     -0.00624    0.04837   -0.13    0.8973
## Year2012      0.02449    0.05153    0.48    0.6347
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.598
## Multiple R-squared:  0.0284, Adjusted R-squared:  0.0251
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 569 weights are ~= 1. The remaining 5912 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0791 0.8670 0.9500 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.54e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.094 1 1.447
## LastAuthorFemale 2.091 1 1.446
## Year 1.023 16 1.001

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32152 -0.40331 0.00858 0.41513 2.33251
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.12842 0.04050 27.87 <2e-16 ***
## FirstAuthorFemale1 0.05858 0.02314 2.53 0.011 *
## LastAuthorFemale1 0.02846 0.02317 1.23 0.219
## Year1997 0.05906 0.05631 1.05 0.294
## Year1998 0.07507 0.05497 1.37 0.172
## Year1999 -0.00636 0.05612 -0.11 0.910
## Year2000 0.09283 0.05431 1.71 0.087 .
## Year2001 0.06150 0.05516 1.11 0.265
## Year2002 0.05430 0.05224 1.04 0.299
## Year2003 0.04388 0.05017 0.87 0.382
## Year2004 0.06597 0.05023 1.31 0.189
## Year2005 0.07089 0.04759 1.49 0.136
```

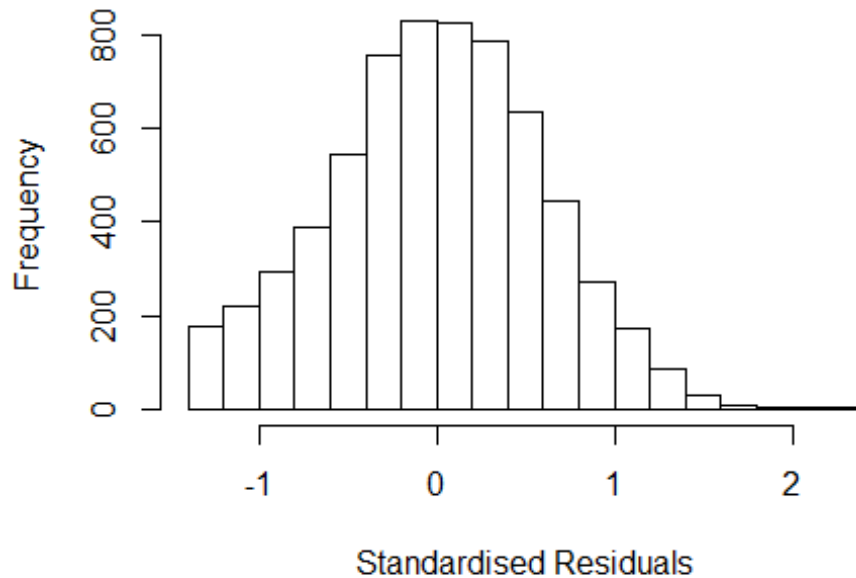


```

## Year2006          0.10607      0.04811      2.20      0.028 *
## Year2007          0.08360      0.05031      1.66      0.097 .
## Year2008          0.04290      0.05092      0.84      0.400
## Year2009          0.04046      0.05008      0.81      0.419
## Year2010          0.08909      0.04967      1.79      0.073 .
## Year2011          0.00893      0.04839      0.18      0.854
## Year2012          0.04320      0.05150      0.84      0.402
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.00673,    Adjusted R-squared:  0.00396
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 562 weights are ~= 1. The remaining 5919 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.106  0.868  0.950  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.015 1          1.008
## Year              1.015 16          1.000

```

## Residuals from first author



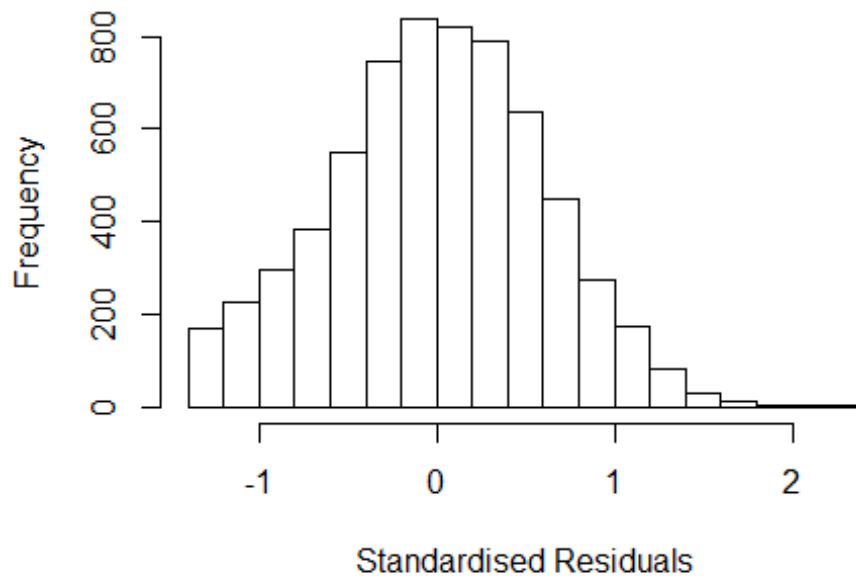
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.31646 -0.40097 0.00697 0.41734 2.33140
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.13097 0.04038 28.01 < 2e-16 ***
## FirstAuthorFemale1 0.07941 0.01610 4.93 8.3e-07 ***
## Year1997 0.05969 0.05629 1.06 0.289
## Year1998 0.07364 0.05489 1.34 0.180
## Year1999 -0.00667 0.05607 -0.12 0.905
## Year2000 0.09232 0.05423 1.70 0.089 .
## Year2001 0.06193 0.05517 1.12 0.262
## Year2002 0.05519 0.05224 1.06 0.291
## Year2003 0.04445 0.05017 0.89 0.376
## Year2004 0.06687 0.05021 1.33 0.183
## Year2005 0.07106 0.04757 1.49 0.135
## Year2006 0.10608 0.04807 2.21 0.027 *
```

```

## Year2007          0.08401    0.05029    1.67    0.095 .
## Year2008          0.04231    0.05090    0.83    0.406
## Year2009          0.04021    0.05005    0.80    0.422
## Year2010          0.08939    0.04963    1.80    0.072 .
## Year2011          0.00925    0.04837    0.19    0.848
## Year2012          0.04336    0.05146    0.84    0.399
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.00651,    Adjusted R-squared:  0.00389
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 564 weights are ~= 1. The remaining 5917 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.106  0.867  0.950  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.014 1      1.007
## Year              1.014 16      1.000

```

## Residuals from last author



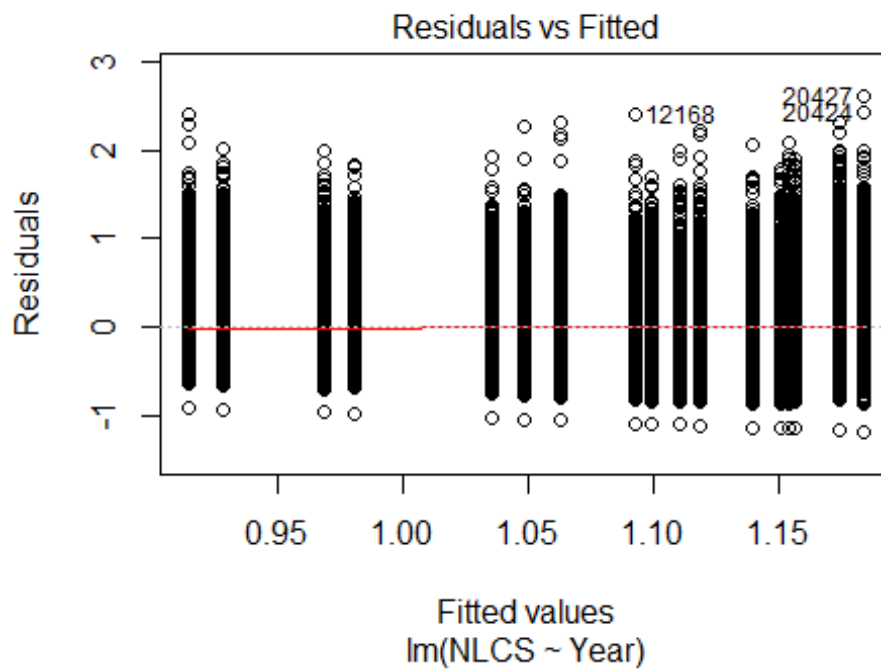
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3127 -0.4034  0.0087  0.4131  2.3261
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.13242    0.04053   27.94 < 2e-16 ***
## LastAuthorFemale1 0.07126    0.01611    4.42 9.9e-06 ***
## Year1997        0.05897    0.05634    1.05  0.295
## Year1998        0.07744    0.05503    1.41  0.159
## Year1999       -0.00479    0.05616   -0.09  0.932
## Year2000        0.09674    0.05444    1.78  0.076 .
## Year2001        0.06241    0.05524    1.13  0.259
## Year2002        0.05468    0.05227    1.05  0.296
## Year2003        0.04295    0.05024    0.85  0.393
## Year2004        0.06608    0.05027    1.31  0.189
## Year2005        0.07147    0.04764    1.50  0.134
## Year2006        0.10906    0.04814    2.27  0.024 *
```

```

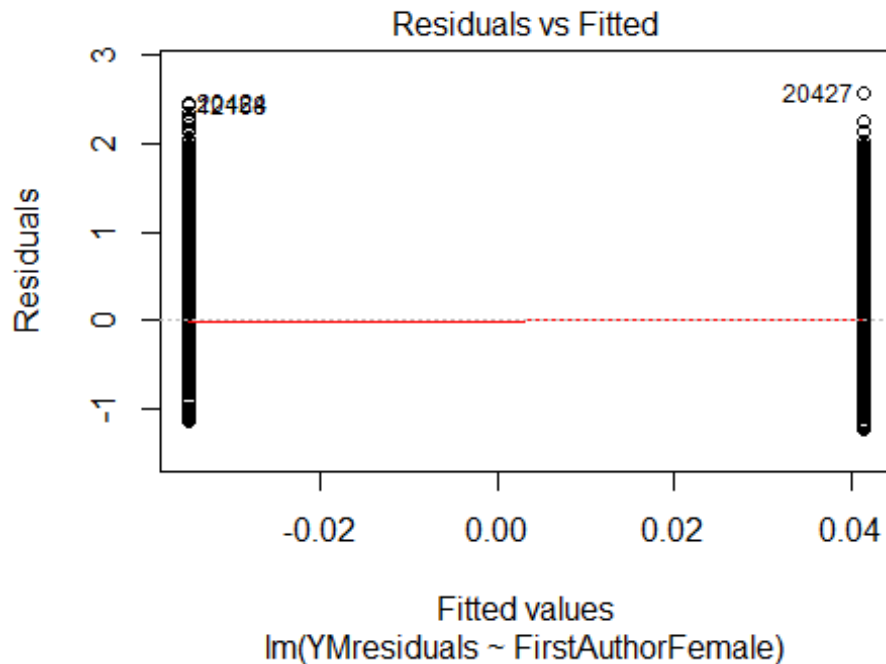
## Year2007      0.08522    0.05032    1.69    0.090 .
## Year2008      0.04489    0.05101    0.88    0.379
## Year2009      0.04211    0.05014    0.84    0.401
## Year2010      0.09134    0.04975    1.84    0.066 .
## Year2011      0.01072    0.04845    0.22    0.825
## Year2012      0.04586    0.05156    0.89    0.374
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.606
## Multiple R-squared:  0.00577,    Adjusted R-squared:  0.00316
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 560 weights are ~= 1. The remaining 5921 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.868  0.950  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.54e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 6481"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3304"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  933 1290 1363 1264 1436 1376 1331 1179 1132 1219 1463 1837 1725 1911 2186
## 2011 2012
## 2287 2345
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  756 1057 1109 1026 1136 1082 1159 1008  977 1031 1247 1586 1458 1609 1853
## 2011 2012

```

```
## 1922 1969
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 731 1024 1059 983 1095 1031 1098 965 926 984 1187 1497 1370 1512 1746
## 2011 2012
## 1804 1885
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

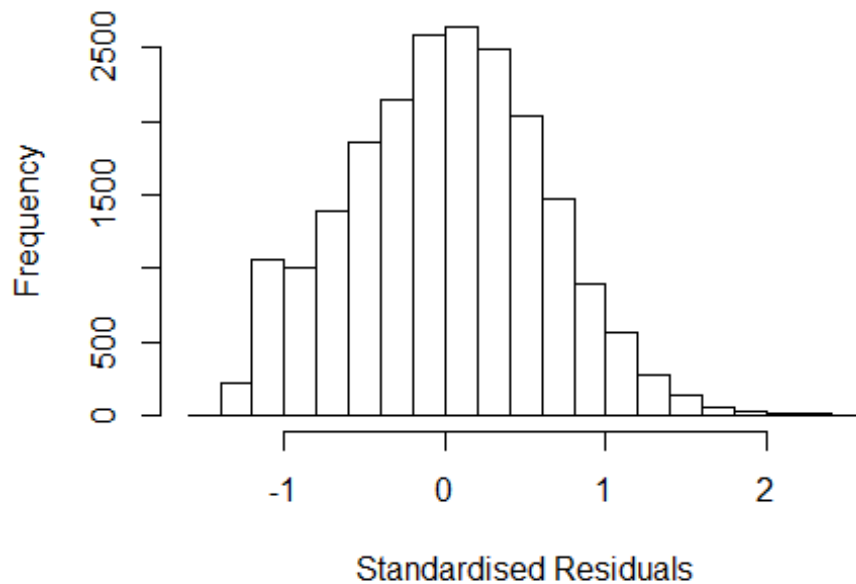


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 19, df = 1, p-value = 1e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 1.8418660596272"
## [1] "Male first author team size 2018 geometric mean: 1.60325005273746"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 380000, p-value = 8e-07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.7389049782506"
## [1] "Male last author team size 2018 geometric mean: 1.72828891315547"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 350000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.458 1      1.208
## LastAuthorFemale  1.450 1      1.204
## UniqueAuthors     1.043 4      1.005
## Year               1.049 16     1.001
```

## Residuals from first and last author and team size



```
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 3020      0032285568 3.323 1998      3304      1      2.527
## 12168     3142680761 3.502 2004      1710      3      2.536
## 20424     72849107558 3.615 2009      2214      2      2.564
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4011 -0.4299  0.0127  0.4228  2.5639
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8734     0.0259   33.67 < 2e-16 ***
## FirstAuthorFemale1 0.0403     0.0106    3.80 0.00015 ***
## LastAuthorFemale1 0.0368     0.0106    3.47 0.00052 ***
## UniqueAuthors2    0.1528     0.0105   14.61 < 2e-16 ***
## UniqueAuthors3    0.2010     0.0141   14.27 < 2e-16 ***
## UniqueAuthors4    0.2525     0.0208   12.13 < 2e-16 ***
## UniqueAuthors5    0.2908     0.0216   13.44 < 2e-16 ***
## Year1997        -0.0523     0.0334   -1.57 0.11720
```

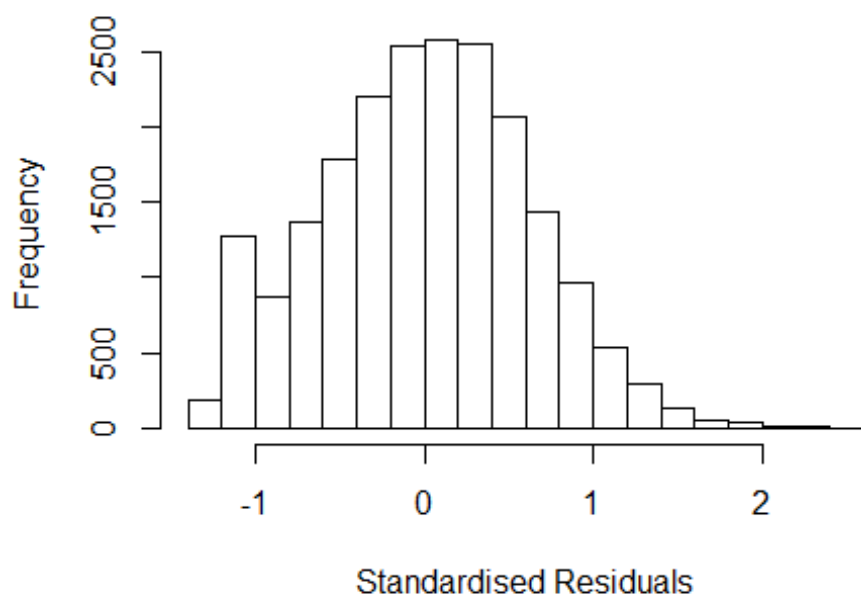


```

## Year1998          -0.0774      0.0321    -2.41  0.01578 *
## Year1999          -0.0204      0.0328    -0.62  0.53391
## Year2000           0.0685      0.0325     2.11  0.03513 *
## Year2001           0.0578      0.0327     1.77  0.07719 .
## Year2002           0.1063      0.0318     3.35  0.00082 ***
## Year2003           0.0621      0.0329     1.89  0.05889 .
## Year2004           0.0929      0.0326     2.85  0.00433 **
## Year2005           0.1052      0.0323     3.26  0.00112 **
## Year2006           0.1121      0.0306     3.66  0.00025 ***
## Year2007           0.1343      0.0293     4.58  4.7e-06 ***
## Year2008           0.1598      0.0304     5.25  1.5e-07 ***
## Year2009           0.1777      0.0303     5.88  4.3e-09 ***
## Year2010           0.1449      0.0296     4.90  9.8e-07 ***
## Year2011           0.1552      0.0297     5.22  1.8e-07 ***
## Year2012           0.1569      0.0307     5.10  3.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0453, Adjusted R-squared:  0.0443
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 1738 weights are ~= 1. The remaining 19159 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0571 0.8710 0.9500 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      4.79e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.468 1 1.211
## LastAuthorFemale 1.466 1 1.211
## Year 1.019 16 1.001

```

## Residuals from first and last author



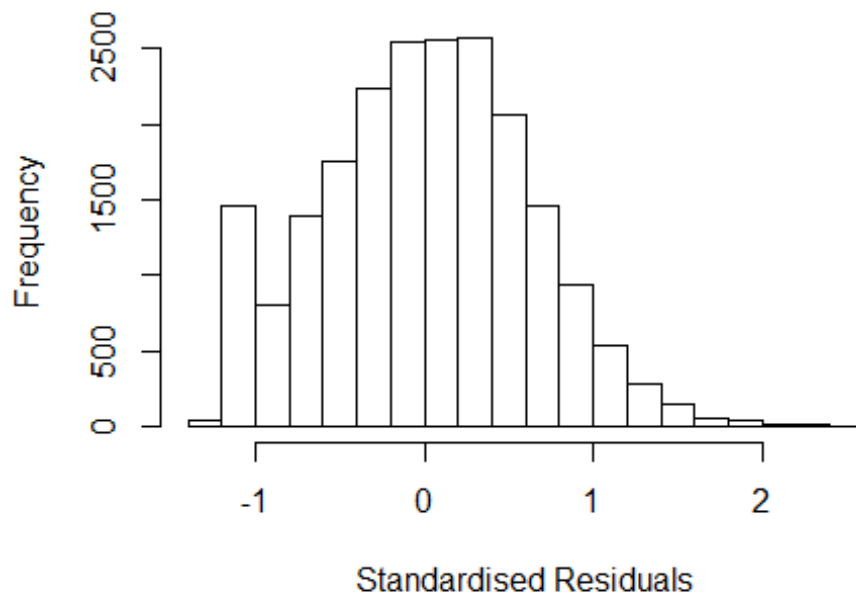
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2262 -0.4274 0.0198 0.4256 2.4815
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9322 0.0264 35.34 < 2e-16 ***
## FirstAuthorFemale1 0.0536 0.0108 4.97 6.8e-07 ***
## LastAuthorFemale1 0.0390 0.0108 3.61 0.00030 ***
## Year1997 -0.0574 0.0341 -1.68 0.09215 .
## Year1998 -0.0744 0.0328 -2.27 0.02334 *
## Year1999 -0.0209 0.0333 -0.63 0.52925
## Year2000 0.0752 0.0331 2.27 0.02307 *
## Year2001 0.0663 0.0333 1.99 0.04614 *
## Year2002 0.1220 0.0325 3.75 0.00018 ***
## Year2003 0.0758 0.0337 2.25 0.02433 *
## Year2004 0.1104 0.0332 3.32 0.00089 ***
## Year2005 0.1240 0.0328 3.78 0.00016 ***
```

```

## Year2006          0.1324      0.0313      4.23  2.4e-05 ***
## Year2007          0.1595      0.0300      5.31  1.1e-07 ***
## Year2008          0.1793      0.0311      5.77  8.1e-09 ***
## Year2009          0.2013      0.0309      6.52  7.2e-11 ***
## Year2010          0.1712      0.0303      5.66  1.5e-08 ***
## Year2011          0.1760      0.0304      5.80  6.7e-09 ***
## Year2012          0.1812      0.0313      5.79  7.3e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.634
## Multiple R-squared:  0.0227, Adjusted R-squared:  0.0218
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 1724 weights are ~= 1. The remaining 19173 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0911 0.8720 0.9500 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.79e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.014 1          1.007
## Year              1.014 16          1.000

```

## Residuals from first author



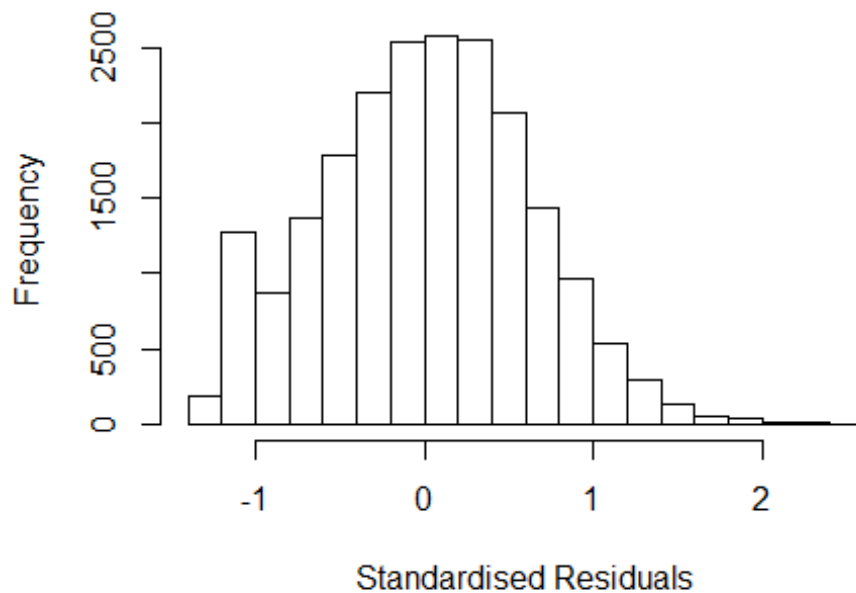
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2176 -0.4236 0.0197 0.4245 2.4739
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93698 0.02632 35.60 < 2e-16 ***
## FirstAuthorFemale1 0.07657 0.00897 8.54 < 2e-16 ***
## Year1997 -0.05672 0.03407 -1.66 0.09603 .
## Year1998 -0.07351 0.03278 -2.24 0.02495 *
## Year1999 -0.01908 0.03324 -0.57 0.56601
## Year2000 0.07642 0.03307 2.31 0.02085 *
## Year2001 0.06756 0.03328 2.03 0.04239 *
## Year2002 0.12480 0.03248 3.84 0.00012 ***
## Year2003 0.07841 0.03367 2.33 0.01988 *
## Year2004 0.11289 0.03320 3.40 0.00067 ***
## Year2005 0.12636 0.03281 3.85 0.00012 ***
## Year2006 0.13474 0.03126 4.31 1.6e-05 ***
```

```

## Year2007          0.16173      0.03002      5.39  7.2e-08 ***
## Year2008          0.18151      0.03106      5.84  5.2e-09 ***
## Year2009          0.20409      0.03084      6.62  3.7e-11 ***
## Year2010          0.17378      0.03023      5.75  9.2e-09 ***
## Year2011          0.17848      0.03034      5.88  4.1e-09 ***
## Year2012          0.18394      0.03129      5.88  4.2e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.634
## Multiple R-squared:  0.0221, Adjusted R-squared:  0.0213
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 1713 weights are ~= 1. The remaining 19184 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0937 0.8720 0.9500 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.79e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.013 1      1.006
## Year      1.013 16      1.000

```

## Residuals from last author



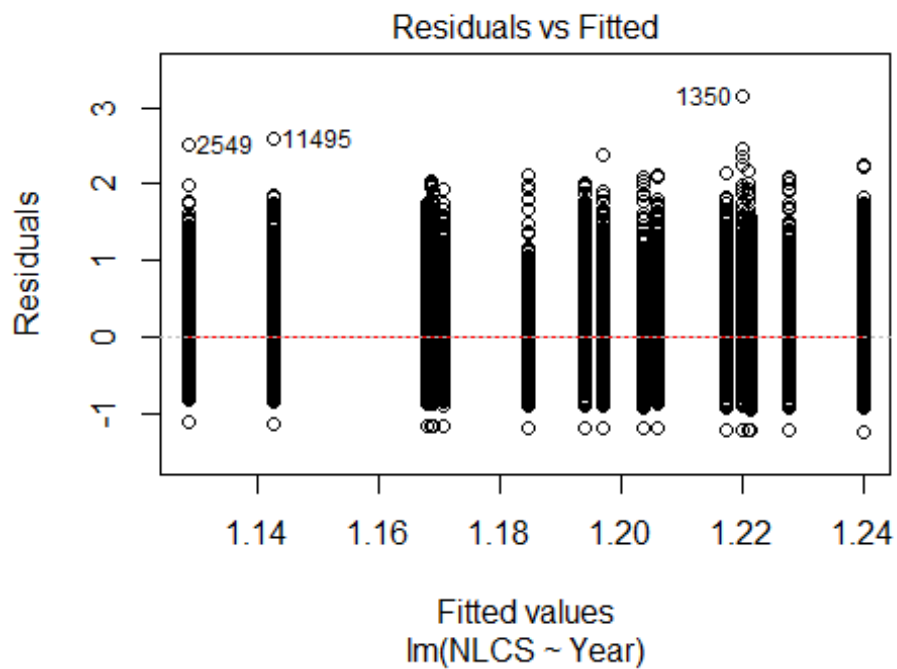
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2150 -0.4267 0.0183 0.4250 2.4707
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.93981 0.02635 35.66 < 2e-16 ***
## LastAuthorFemale1 0.07071 0.00899 7.87 3.9e-15 ***
## Year1997 -0.05576 0.03410 -1.63 0.10207
## Year1998 -0.07272 0.03280 -2.22 0.02665 *
## Year1999 -0.01993 0.03328 -0.60 0.54929
## Year2000 0.07704 0.03306 2.33 0.01978 *
## Year2001 0.06888 0.03326 2.07 0.03836 *
## Year2002 0.12372 0.03253 3.80 0.00014 ***
## Year2003 0.07724 0.03366 2.30 0.02174 *
## Year2004 0.11339 0.03323 3.41 0.00064 ***
## Year2005 0.12664 0.03283 3.86 0.00011 ***
## Year2006 0.13607 0.03134 4.34 1.4e-05 ***
```

```

## Year2007      0.16319      0.03005      5.43  5.7e-08 ***
## Year2008      0.18315      0.03108      5.89  3.9e-09 ***
## Year2009      0.20449      0.03089      6.62  3.7e-11 ***
## Year2010      0.17491      0.03027      5.78  7.6e-09 ***
## Year2011      0.18057      0.03035      5.95  2.7e-09 ***
## Year2012      0.18592      0.03132      5.94  3.0e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.634
## Multiple R-squared:  0.0215, Adjusted R-squared:  0.0207
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 1689 weights are ~= 1. The remaining 19208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0951 0.8720 0.9500 0.9120 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.79e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 20897"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3305"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 951 976 1153 1194 1356 1297 1312 1369 1410 1290 1442 1454 1381 1512 1609
## 2011 2012
## 1690 1735
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 675 678 715 714 765 507 1082 1165 1234 1113 1226 1242 1160 1254 1346
## 2011 2012

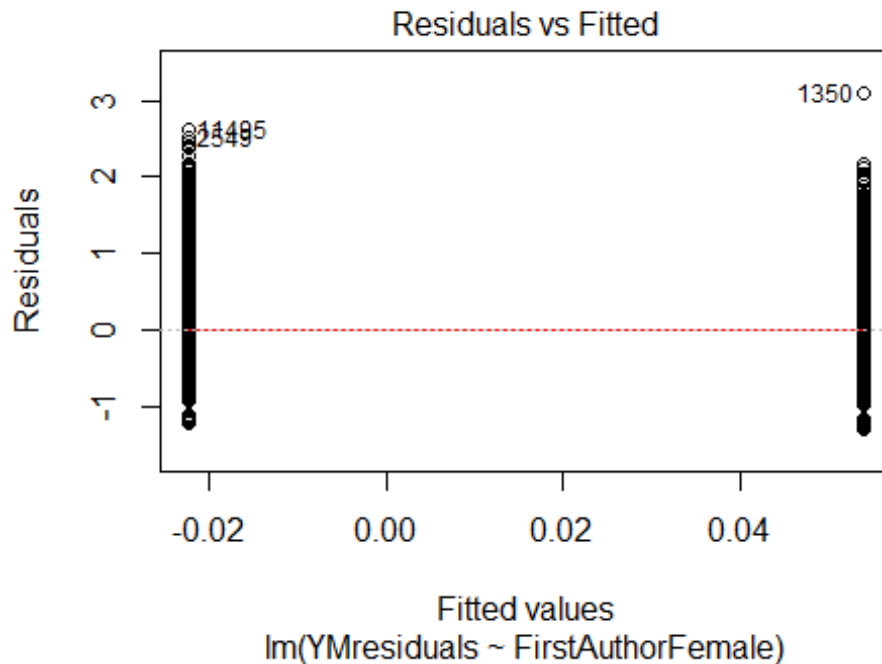
```

```
## 1436 1442
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 651 665 700 694 739 487 1057 1142 1201 1081 1191 1192 1119 1217 1292
## 2011 2012
## 1378 1375
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 9e-04
```



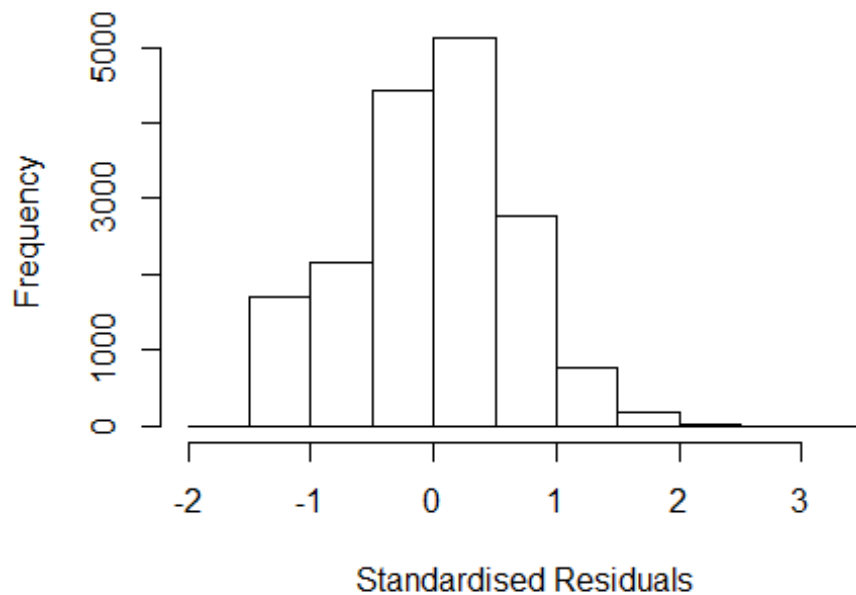
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 7.2, df = 1, p-value = 0.007
```





```
## [1] "Female first author team size 2018 geometric mean: 1.55231580780247"
## [1] "Male first author team size 2018 geometric mean: 1.45532083307372"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190000, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.50204453374793"
## [1] "Male last author team size 2018 geometric mean: 1.48495486752839"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 170000, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.887 1          1.374
## LastAuthorFemale  1.897 1          1.377
## UniqueAuthors    1.030 4          1.004
## Year              1.045 16         1.001
```

## Residuals from first and last author and team size



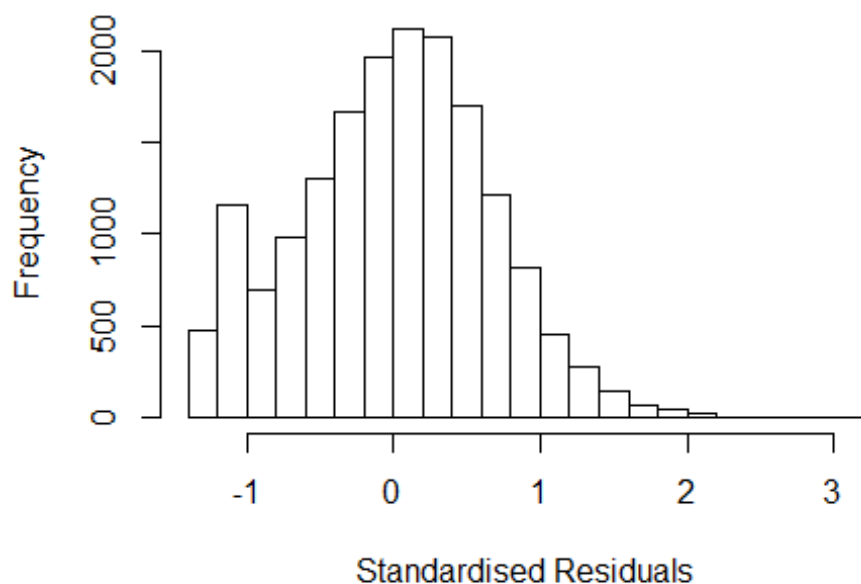
```
## [1] "List of 4 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 1350    0031404342 4.363 1997    3305      1    3.147
## 1362    0031414123 3.668 1997    3305      1    2.543
## 2549    0031877127 3.627 1998    3305      1    2.597
## 11495   11944262160 3.728 2004    3305      1    2.681
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5454 -0.4380  0.0272  0.4413  3.1471
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.11747    0.02982   37.47 < 2e-16 ***
## FirstAuthorFemale1 0.05275    0.01534    3.44 0.00059 ***
## LastAuthorFemale1 0.03836    0.01558    2.46 0.01381 *
## UniqueAuthors2    0.18885    0.01288   14.66 < 2e-16 ***
## UniqueAuthors3    0.26525    0.01861   14.25 < 2e-16 ***
## UniqueAuthors4    0.21020    0.03166    6.64 3.2e-11 ***
## UniqueAuthors5    0.32946    0.03471    9.49 < 2e-16 ***
```

```

## Year1997      0.00735    0.04069    0.18  0.85666
## Year1998     -0.08742    0.04020   -2.17  0.02968 *
## Year1999     -0.00713    0.03895   -0.18  0.85472
## Year2000      0.01715    0.03802    0.45  0.65185
## Year2001     -0.04145    0.04346   -0.95  0.34023
## Year2002     -0.02974    0.03605   -0.82  0.40947
## Year2003     -0.03977    0.03599   -1.11  0.26915
## Year2004     -0.07011    0.03530   -1.99  0.04701 *
## Year2005     -0.03725    0.03602   -1.03  0.30110
## Year2006      0.01127    0.03505    0.32  0.74771
## Year2007      0.00510    0.03579    0.14  0.88666
## Year2008     -0.02247    0.03668   -0.61  0.54005
## Year2009     -0.03569    0.03505   -1.02  0.30859
## Year2010      0.00348    0.03469    0.10  0.92003
## Year2011     -0.02559    0.03453   -0.74  0.45876
## Year2012      0.01115    0.03472    0.32  0.74818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.028, Adjusted R-squared:  0.0267
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 864 is an outlier with |weight| = 0 ( < 5.8e-06);
## 1482 weights are ~= 1. The remaining 15698 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0514 0.8660 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.82e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.958 1          1.399
## LastAuthorFemale 1.953 1          1.398
## Year 1.021 16          1.001

```

## Residuals from first and last author



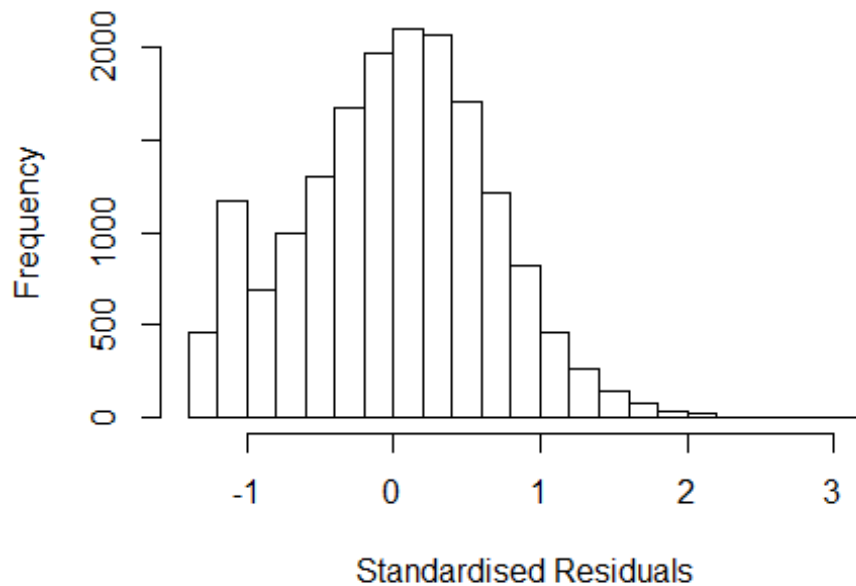
```
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 1350    0031404342 4.363 1997    3305      1    3.094
## 2549    0031877127 3.627 1998    3305      1    2.533
## 11495 11944262160 3.728 2004    3305      1    2.618
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2905 -0.4471  0.0318  0.4489  3.0938
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.16862    0.02976   39.27 < 2e-16 ***
## FirstAuthorFemale1 0.05709    0.01585    3.60 0.00032 ***
## LastAuthorFemale1 0.02981    0.01604    1.86 0.06318 .
## Year1997         0.01364    0.04090    0.33 0.73870
## Year1998        -0.07461    0.04055   -1.84 0.06577 .
## Year1999         0.00232    0.03928    0.06 0.95298
## Year2000         0.03503    0.03817    0.92 0.35884
## Year2001        -0.02153    0.04371   -0.49 0.62233
## Year2002        -0.02166    0.03617   -0.60 0.54920
## Year2003        -0.03073    0.03622   -0.85 0.39630
```

```

## Year2004          -0.05873      0.03556      -1.65      0.09864 .
## Year2005          -0.02455      0.03628      -0.68      0.49855
## Year2006           0.02499      0.03522       0.71      0.47794
## Year2007           0.02040      0.03597       0.57      0.57060
## Year2008          -0.00372      0.03693      -0.10      0.91984
## Year2009          -0.01541      0.03526      -0.44      0.66208
## Year2010           0.02547      0.03475       0.73      0.46355
## Year2011          -0.00427      0.03462      -0.12      0.90185
## Year2012           0.03441      0.03477       0.99      0.32238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.00529,    Adjusted R-squared:  0.00425
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 864 is an outlier with |weight| = 0 ( < 5.8e-06);
## 1403 weights are ~= 1. The remaining 15777 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0805  0.8650  0.9500  0.9070  0.9860  0.9990
## Algorithmic parameters:
##           tuning.chi             bb           tuning.psi           refine.tol
##           1.55e+00             5.00e-01           4.69e+00           1.00e-07
##           rel.tol             solve.tol           eps.outlier           eps.x
##           1.00e-07             1.00e-07           5.82e-06           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01             5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi             subsampling             cov
##           "bisquare"             "nonsingular"             ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.017 1           1.008
## Year              1.017 16           1.001

```

## Residuals from first author



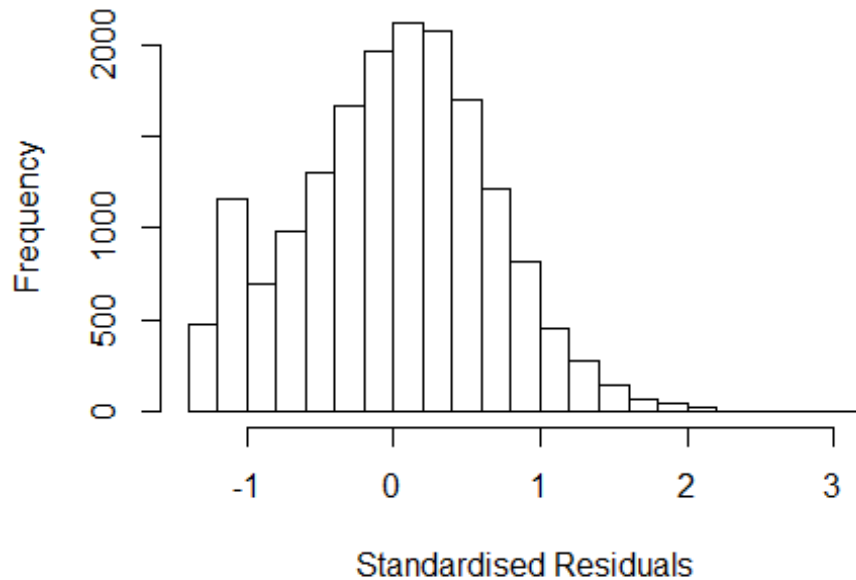
```
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 1350    0031404342 4.363 1997    3305      1    3.094
## 2549    0031877127 3.627 1998    3305      1    2.533
## 11495  11944262160 3.728 2004    3305      1    2.618
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2848 -0.4469  0.0314  0.4484  3.1004
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17094    0.02972   39.40 < 2e-16 ***
## FirstAuthorFemale1 0.07867    0.01143    6.89 5.9e-12 ***
## Year1997         0.01297    0.04089    0.32  0.751
## Year1998        -0.07555    0.04054   -1.86  0.062 .
## Year1999         0.00201    0.03929    0.05  0.959
## Year2000         0.03453    0.03816    0.90  0.366
## Year2001        -0.02140    0.04369   -0.49  0.624
## Year2002        -0.02131    0.03618   -0.59  0.556
## Year2003        -0.03116    0.03622   -0.86  0.390
## Year2004        -0.05899    0.03557   -1.66  0.097 .
```

```

## Year2005      -0.02474    0.03629   -0.68    0.495
## Year2006      0.02493    0.03524    0.71    0.479
## Year2007      0.02021    0.03600    0.56    0.574
## Year2008     -0.00337    0.03695   -0.09    0.927
## Year2009     -0.01502    0.03527   -0.43    0.670
## Year2010      0.02541    0.03476    0.73    0.465
## Year2011     -0.00399    0.03463   -0.12    0.908
## Year2012      0.03513    0.03479    1.01    0.313
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.00509,    Adjusted R-squared:  0.00411
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 864 is an outlier with |weight| = 0 ( < 5.8e-06);
## 1404 weights are ~= 1. The remaining 15776 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0813 0.8650 0.9500 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.82e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.015 1          1.007
## Year            1.015 16          1.000

```

## Residuals from last author



```
## [1] "List of 3 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 1350    0031404342 4.363 1997    3305      1    3.094
## 2549    0031877127 3.627 1998    3305      1    2.533
## 11495  11944262160 3.728 2004    3305      1    2.618
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2797 -0.4459  0.0312  0.4487  3.1050
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17178    0.02971   39.44 < 2e-16 ***
## LastAuthorFemale1 0.07147    0.01155    6.19 6.3e-10 ***
## Year1997        0.01474    0.04089    0.36  0.718
## Year1998       -0.07386    0.04052   -1.82  0.068 .
## Year1999        0.00249    0.03927    0.06  0.950
## Year2000        0.03622    0.03820    0.95  0.343
## Year2001       -0.02118    0.04374   -0.48  0.628
## Year2002       -0.02156    0.03614   -0.60  0.551
## Year2003       -0.02988    0.03622   -0.83  0.409
## Year2004       -0.05716    0.03554   -1.61  0.108
```

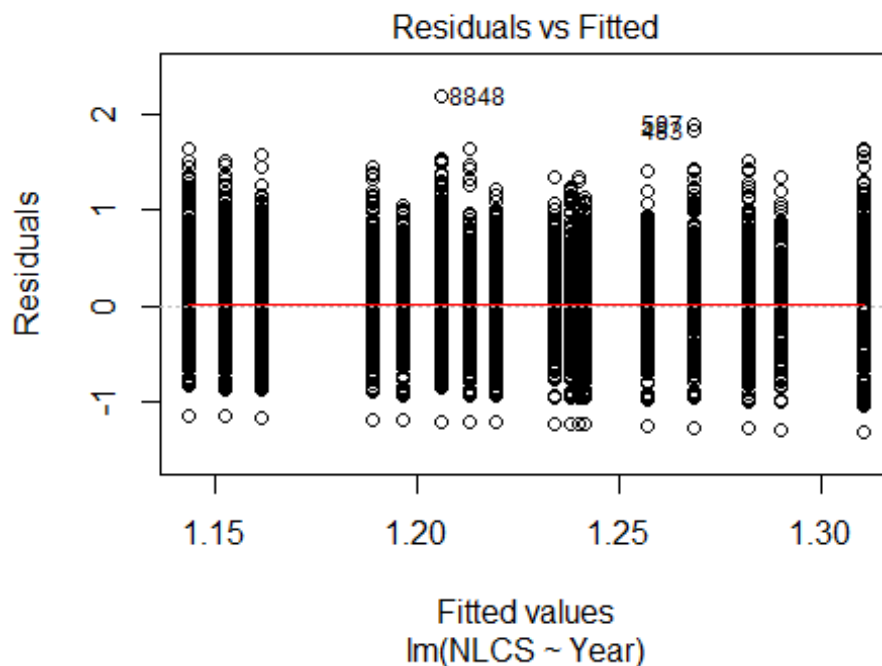


```

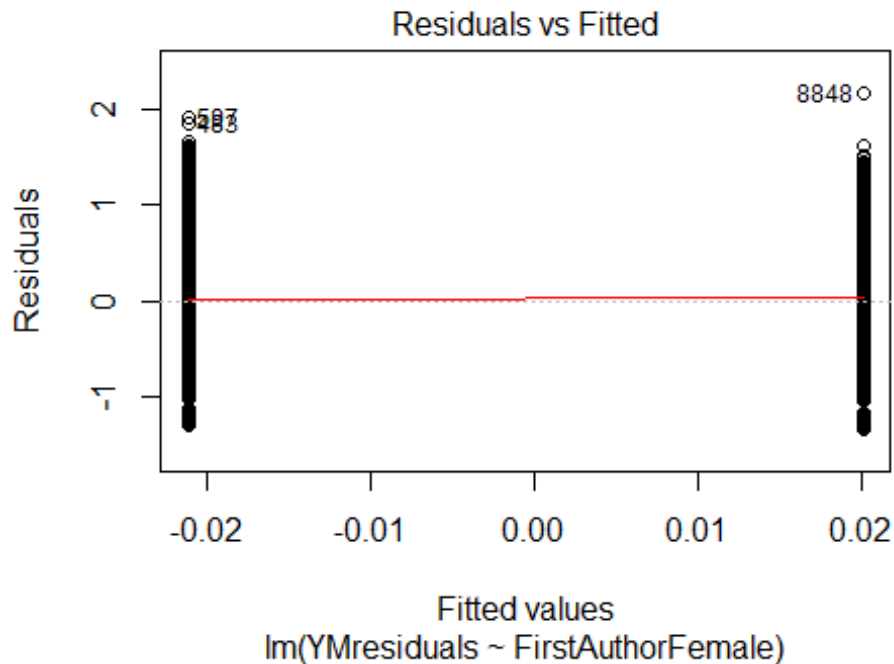
## Year2005          -0.02352      0.03626      -0.65      0.517
## Year2006           0.02790      0.03516       0.79      0.427
## Year2007           0.02266      0.03593       0.63      0.528
## Year2008          -0.00194      0.03689      -0.05      0.958
## Year2009          -0.01366      0.03523      -0.39      0.698
## Year2010           0.02814      0.03471       0.81      0.418
## Year2011          -0.00164      0.03457      -0.05      0.962
## Year2012           0.03646      0.03472       1.05      0.294
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.00459,    Adjusted R-squared:  0.0036
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 864 is an outlier with |weight| = 0 ( < 5.8e-06);
## 1404 weights are ~= 1. The remaining 15776 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0822 0.8640 0.9500 0.9070 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      5.82e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 17181"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3306"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 271 310 337 340 426 391 401 334 302 395 469 606 542 597 733
## 2011 2012
## 702 681
##

```

```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 200 218 256 263 299 259 334 279 250 325 385 510 447 494 615
## 2011 2012
## 611 578
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 195 211 250 253 286 252 316 265 230 298 361 482 415 467 586
## 2011 2012
## 572 549
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

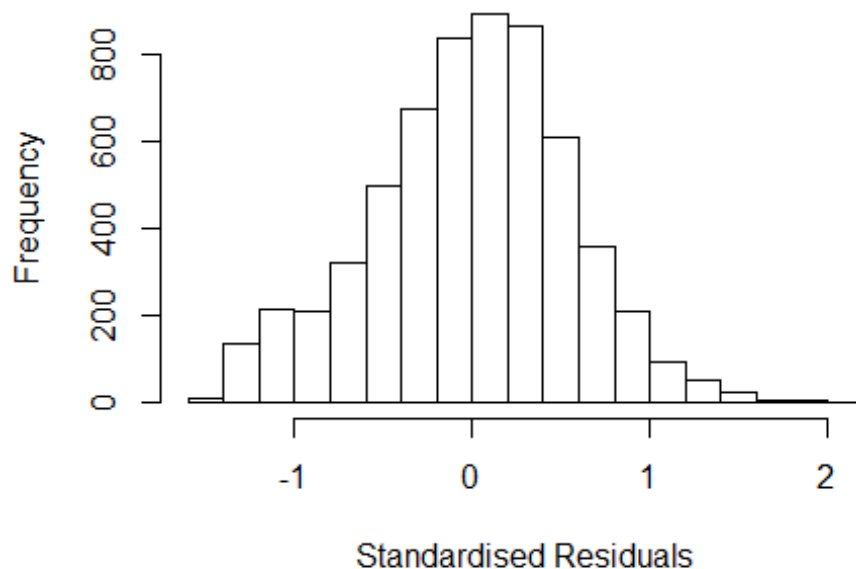


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00012, df = 1, p-value = 1
```



```
## [1] "Female first author team size 2018 geometric mean: 2.12489709243516"
## [1] "Male first author team size 2018 geometric mean: 2.05812396112557"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 48000, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.07771428043003"
## [1] "Male last author team size 2018 geometric mean: 2.12746818386801"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 47000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.377 1          1.173
## LastAuthorFemale  1.366 1          1.169
## UniqueAuthors     1.093 4          1.011
## Year               1.095 16         1.003
```

## Residuals from first and last author and team size



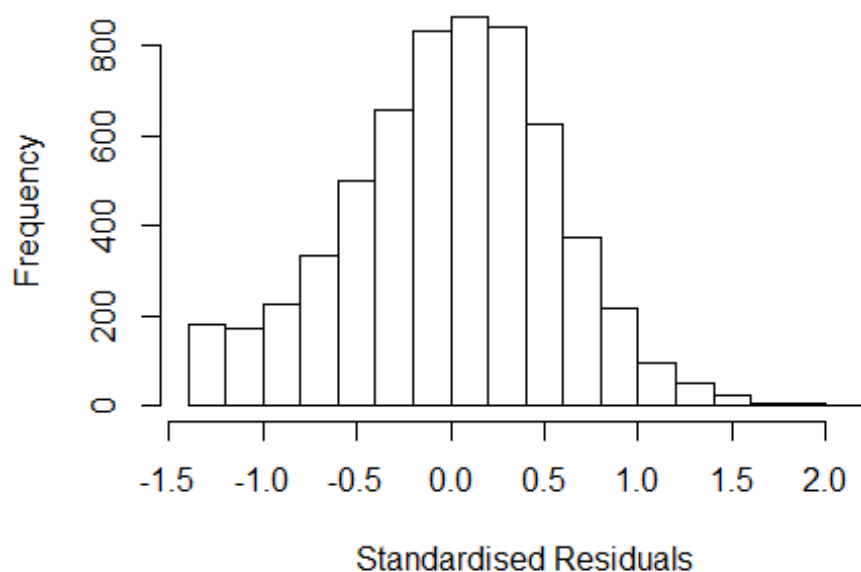
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4664 -0.3624 0.0205 0.3598 2.1050
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2295 0.0569 21.60 < 2e-16 ***
## FirstAuthorFemale1 0.0273 0.0173 1.58 0.11481
## LastAuthorFemale1 0.0252 0.0173 1.46 0.14451
## UniqueAuthors2 0.1209 0.0190 6.38 1.9e-10 ***
## UniqueAuthors3 0.1844 0.0220 8.39 < 2e-16 ***
## UniqueAuthors4 0.2534 0.0282 8.97 < 2e-16 ***
## UniqueAuthors5 0.2100 0.0269 7.81 6.5e-15 ***
## Year1997 -0.0673 0.0748 -0.90 0.36848
## Year1998 -0.1369 0.0707 -1.94 0.05294 .
## Year1999 -0.0690 0.0665 -1.04 0.29921
```

```

## Year2000          -0.0149      0.0655   -0.23   0.81962
## Year2001          -0.0417      0.0690   -0.61   0.54500
## Year2002          -0.0765      0.0641   -1.19   0.23271
## Year2003          -0.0831      0.0633   -1.31   0.18946
## Year2004          -0.0811      0.0658   -1.23   0.21770
## Year2005          -0.1302      0.0635   -2.05   0.04043 *
## Year2006          -0.1304      0.0626   -2.08   0.03718 *
## Year2007          -0.1145      0.0612   -1.87   0.06135 .
## Year2008          -0.1076      0.0610   -1.76   0.07789 .
## Year2009          -0.1750      0.0612   -2.86   0.00424 **
## Year2010          -0.1736      0.0608   -2.85   0.00434 **
## Year2011          -0.2044      0.0620   -3.30   0.00098 ***
## Year2012          -0.1511      0.0628   -2.41   0.01611 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.0354, Adjusted R-squared:  0.0318
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 501 weights are ~= 1. The remaining 5487 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0886 0.8640 0.9510 0.9010 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.67e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.387 1          1.178
## LastAuthorFemale 1.386 1          1.177
## Year              1.033 16          1.001

```

## Residuals from first and last author



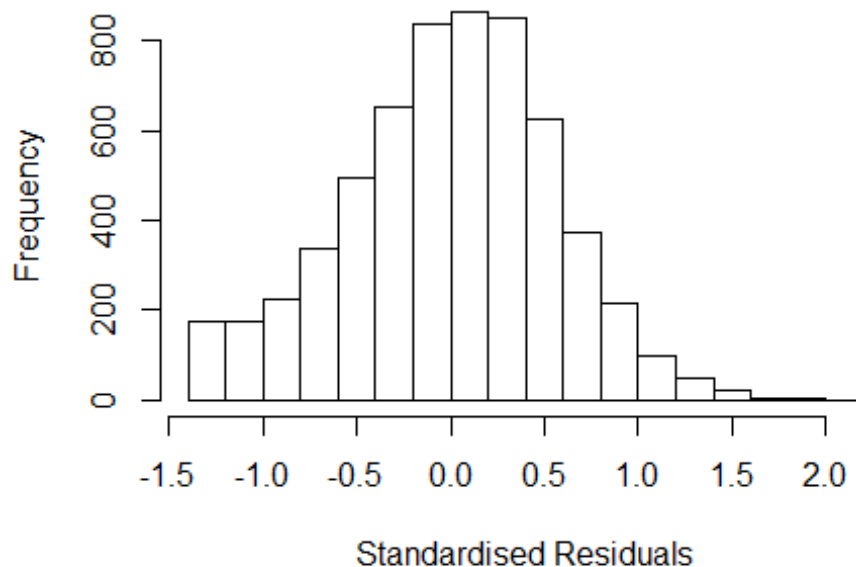
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3479 -0.3664 0.0267 0.3677 2.1802
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.287029 0.058537 21.99 <2e-16 ***
## FirstAuthorFemale1 0.049268 0.017663 2.79 0.0053 **
## LastAuthorFemale1 0.011527 0.017682 0.65 0.5145
## Year1997 -0.062037 0.076926 -0.81 0.4200
## Year1998 -0.119776 0.073488 -1.63 0.1032
## Year1999 -0.074653 0.068306 -1.09 0.2745
## Year2000 0.000102 0.067060 0.00 0.9988
## Year2001 -0.021694 0.070820 -0.31 0.7594
## Year2002 -0.061889 0.065614 -0.94 0.3456
## Year2003 -0.062777 0.065311 -0.96 0.3365
## Year2004 -0.048622 0.068339 -0.71 0.4768
## Year2005 -0.098054 0.065601 -1.49 0.1350
```

```

## Year2006      -0.095846    0.064391   -1.49    0.1367
## Year2007      -0.083848    0.063062   -1.33    0.1837
## Year2008      -0.070565    0.062921   -1.12    0.2621
## Year2009      -0.151667    0.063240   -2.40    0.0165 *
## Year2010      -0.148486    0.062726   -2.37    0.0180 *
## Year2011      -0.177250    0.063963   -2.77    0.0056 **
## Year2012      -0.121523    0.064580   -1.88    0.0599 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.00987,    Adjusted R-squared:  0.00689
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 473 weights are ~= 1. The remaining 5515 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0719 0.8640 0.9500 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.018 1      1.009
## Year      1.018 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3449 -0.3664 0.0253 0.3670 2.1714
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29e+00 5.84e-02 22.08 < 2e-16 ***
## FirstAuthorFemale1 5.55e-02 1.51e-02 3.67 0.00025 ***
## Year1997 -6.26e-02 7.69e-02 -0.81 0.41587
## Year1998 -1.20e-01 7.35e-02 -1.63 0.10235
## Year1999 -7.50e-02 6.83e-02 -1.10 0.27230
## Year2000 9.73e-05 6.70e-02 0.00 0.99884
## Year2001 -2.20e-02 7.08e-02 -0.31 0.75546
## Year2002 -6.16e-02 6.56e-02 -0.94 0.34775
## Year2003 -6.30e-02 6.53e-02 -0.96 0.33464
## Year2004 -4.83e-02 6.83e-02 -0.71 0.47971
## Year2005 -9.79e-02 6.56e-02 -1.49 0.13547
## Year2006 -9.60e-02 6.44e-02 -1.49 0.13584
```

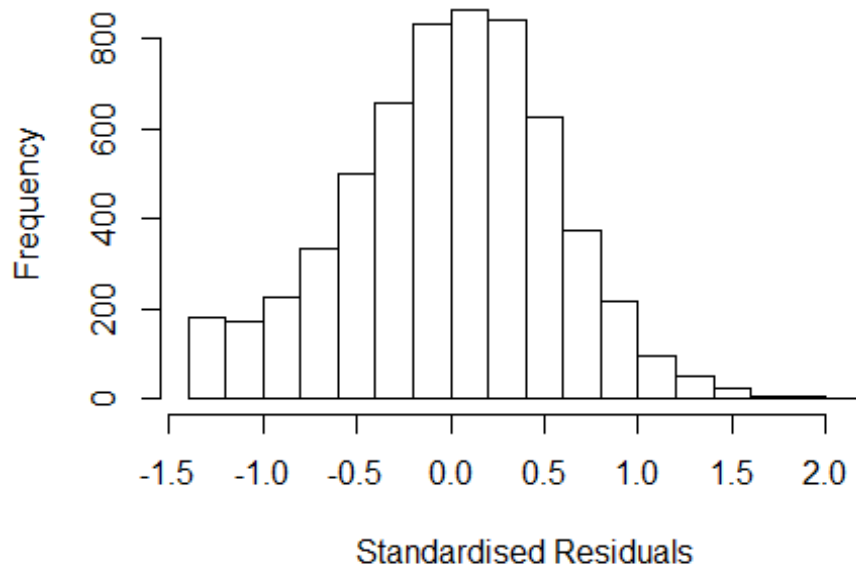


```

## Year2007          -8.43e-02   6.30e-02   -1.34   0.18098
## Year2008          -7.02e-02   6.29e-02   -1.12   0.26444
## Year2009          -1.52e-01   6.32e-02   -2.40   0.01634 *
## Year2010          -1.48e-01   6.27e-02   -2.36   0.01824 *
## Year2011          -1.78e-01   6.39e-02   -2.78   0.00550 **
## Year2012          -1.21e-01   6.46e-02   -1.88   0.06041 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.00979,    Adjusted R-squared:  0.00697
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 478 weights are ~= 1. The remaining 5510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0757 0.8640 0.9500 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1          1.009
## Year            1.017 16          1.001

```

## Residuals from last author



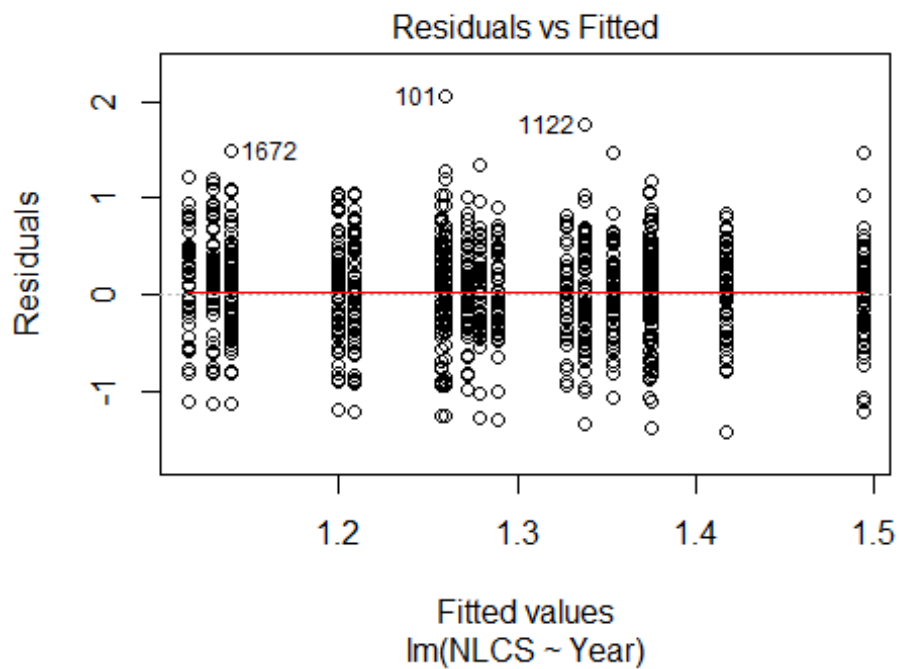
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3366 -0.3681 0.0248 0.3702 2.2154
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.29563 0.05865 22.09 <2e-16 ***
## LastAuthorFemale1 0.03818 0.01516 2.52 0.0118 *
## Year1997 -0.06034 0.07717 -0.78 0.4343
## Year1998 -0.11791 0.07364 -1.60 0.1094
## Year1999 -0.07167 0.06839 -1.05 0.2947
## Year2000 0.00277 0.06714 0.04 0.9671
## Year2001 -0.01564 0.07087 -0.22 0.8253
## Year2002 -0.05924 0.06577 -0.90 0.3678
## Year2003 -0.05965 0.06541 -0.91 0.3619
## Year2004 -0.04820 0.06852 -0.70 0.4818
## Year2005 -0.09449 0.06573 -1.44 0.1506
## Year2006 -0.09229 0.06451 -1.43 0.1526
```

```

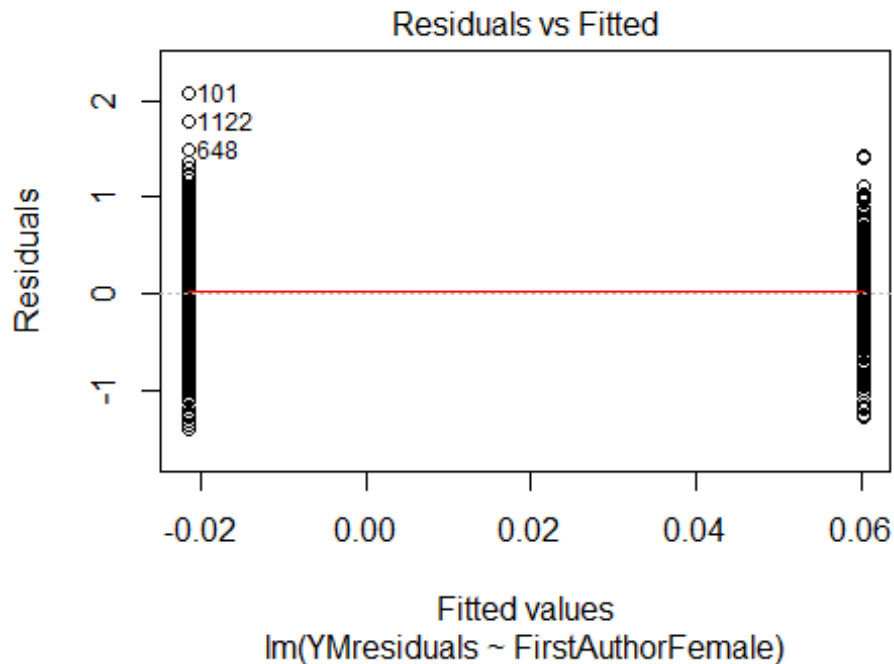
## Year2007          -0.07804      0.06318    -1.24    0.2168
## Year2008          -0.06739      0.06304    -1.07    0.2851
## Year2009          -0.14594      0.06335    -2.30    0.0213 *
## Year2010          -0.14391      0.06285    -2.29    0.0221 *
## Year2011          -0.17175      0.06404    -2.68    0.0073 **
## Year2012          -0.11608      0.06468    -1.79    0.0728 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.0085, Adjusted R-squared:  0.00568
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 466 weights are ~= 1. The remaining 5522 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0612 0.8640 0.9500 0.9020 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.67e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 5988"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3307"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   96   80  103   87  139   94   65   62   86   65   89   92   86   98   74
## 2011 2012
##   88  106
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   73   51   72   63   91   58   46   43   55   47   63   65   59   82   54
## 2011 2012

```

```
## 71 85
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 67 46 66 55 88 54 44 39 52 42 55 62 57 76 50
## 2011 2012
## 68 80
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



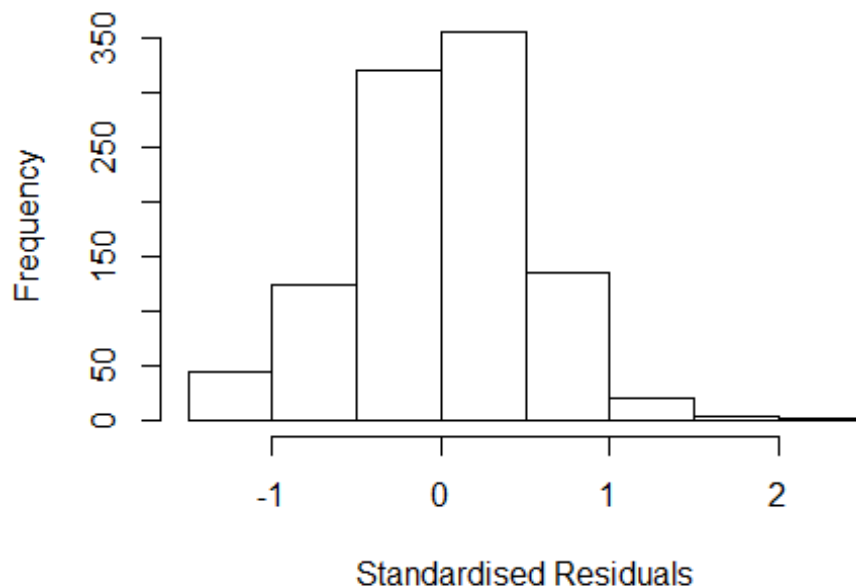
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3.6, df = 1, p-value = 0.06
## [1] "Female first author team size 2018 geometric mean: 2.83692798813327"
## [1] "Male first author team size 2018 geometric mean: 2.49073545770665"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 820, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.70960378087943"
## [1] "Male last author team size 2018 geometric mean: 2.64769107735197"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 570, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.151	1	1.073
LastAuthorFemale	1.120	1	1.058
UniqueAuthors	1.633	4	1.063
Year	1.763	16	1.018

## Residuals from first and last author and team size



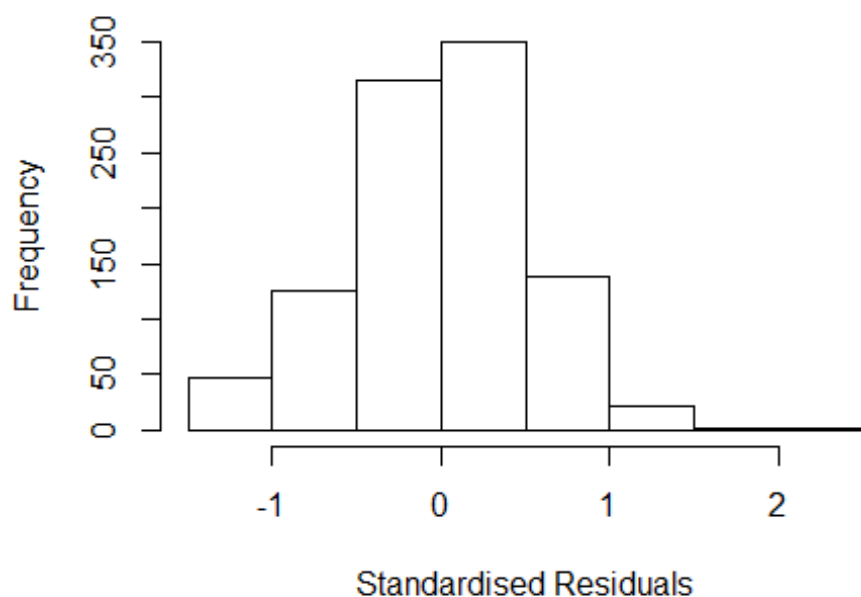
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.404 -0.322 0.013 0.339 2.091
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2080 0.0820 14.74 <2e-16 ***
## FirstAuthorFemale1 0.0703 0.0382 1.84 0.066 .
## LastAuthorFemale1 -0.0384 0.0436 -0.88 0.378
## UniqueAuthors2 0.0541 0.0450 1.20 0.230
## UniqueAuthors3 0.1090 0.0524 2.08 0.038 *
## UniqueAuthors4 0.1196 0.0701 1.71 0.088 .
## UniqueAuthors5 0.1327 0.0742 1.79 0.074 .
## Year1997 -0.1617 0.1268 -1.28 0.202
## Year1998 -0.0674 0.1027 -0.66 0.512
## Year1999 0.0504 0.1153 0.44 0.662
```

```

## Year2000          0.1257      0.0950      1.32      0.186
## Year2001          0.2037      0.1044      1.95      0.051 .
## Year2002          0.0289      0.1077      0.27      0.789
## Year2003          0.1088      0.1104      0.99      0.325
## Year2004         -0.0442      0.1003     -0.44      0.660
## Year2005          0.0331      0.1093      0.30      0.762
## Year2006          0.1255      0.1035      1.21      0.226
## Year2007          0.0547      0.1079      0.51      0.613
## Year2008          0.0558      0.0999      0.56      0.576
## Year2009         -0.0741      0.1078     -0.69      0.492
## Year2010          0.0648      0.1053      0.62      0.538
## Year2011         -0.1167      0.1287     -0.91      0.364
## Year2012         -0.1439      0.0999     -1.44      0.150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0466, Adjusted R-squared:  0.0252
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 81 weights are ~= 1. The remaining 920 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0527 0.8650 0.9520 0.9010 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.99e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.069 1          1.034
## LastAuthorFemale 1.102 1          1.050
## Year              1.153 16          1.004

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.405 -0.324 0.017 0.338 2.112
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2407 0.0831 14.93 <2e-16 ***
## FirstAuthorFemale1 0.0827 0.0369 2.24 0.025 *
## LastAuthorFemale1 -0.0379 0.0434 -0.87 0.383
## Year1997 -0.1496 0.1282 -1.17 0.244
## Year1998 -0.0525 0.1037 -0.51 0.613
## Year1999 0.0429 0.1173 0.37 0.714
## Year2000 0.1397 0.0960 1.45 0.146
## Year2001 0.2249 0.1045 2.15 0.032 *
## Year2002 0.0631 0.1063 0.59 0.553
## Year2003 0.1082 0.1121 0.97 0.335
## Year2004 -0.0131 0.1016 -0.13 0.897
## Year2005 0.0503 0.1102 0.46 0.648
```

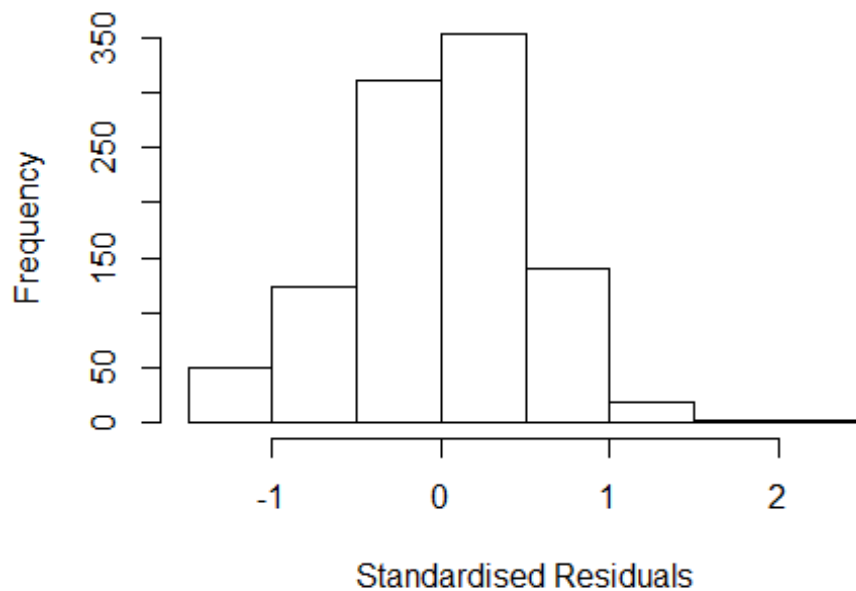


```

## Year2006          0.1640      0.1037      1.58      0.114
## Year2007          0.0897      0.1070      0.84      0.402
## Year2008          0.0879      0.1007      0.87      0.383
## Year2009         -0.0408      0.1088     -0.38      0.708
## Year2010          0.0934      0.1076      0.87      0.386
## Year2011         -0.0812      0.1261     -0.64      0.520
## Year2012         -0.1091      0.1016     -1.07      0.283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0396, Adjusted R-squared:  0.022
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 96 weights are ~= 1. The remaining 905 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.046  0.861  0.953  0.899  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.063 1          1.031
## Year              1.063 16          1.002

```

## Residuals from first author

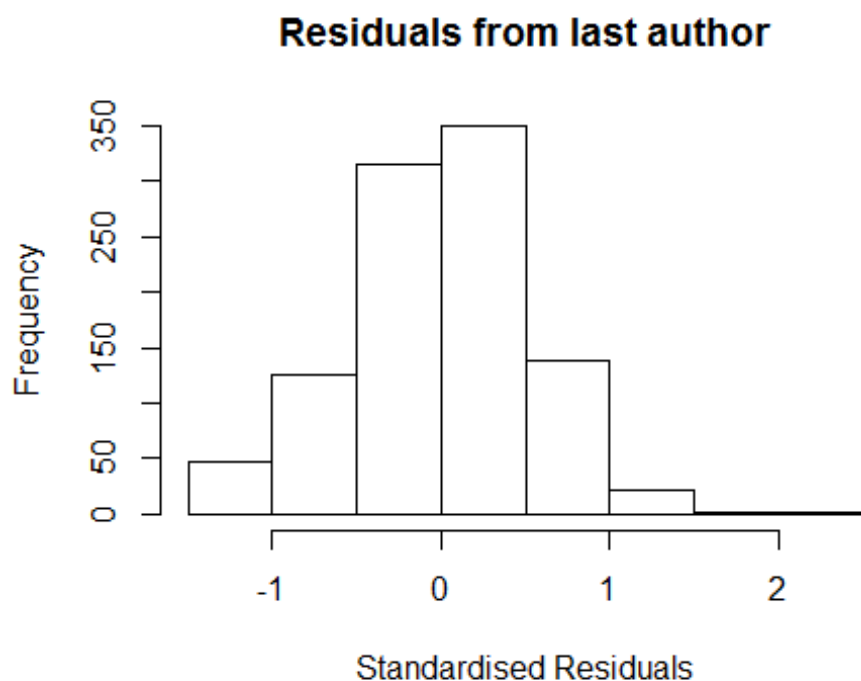


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.399 -0.333 0.021 0.336 2.078
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2366 0.0830 14.89 <2e-16 ***
## FirstAuthorFemale1 0.0733 0.0373 1.96 0.050 *
## Year1997 -0.1453 0.1281 -1.13 0.257
## Year1998 -0.0546 0.1035 -0.53 0.598
## Year1999 0.0398 0.1173 0.34 0.734
## Year2000 0.1386 0.0961 1.44 0.149
## Year2001 0.2216 0.1042 2.13 0.034 *
## Year2002 0.0593 0.1059 0.56 0.576
## Year2003 0.1034 0.1122 0.92 0.357
## Year2004 -0.0166 0.1015 -0.16 0.870
## Year2005 0.0499 0.1101 0.45 0.651
## Year2006 0.1626 0.1037 1.57 0.117
```

```

## Year2007          0.0857      0.1066      0.80      0.422
## Year2008          0.0887      0.1010      0.88      0.380
## Year2009         -0.0444      0.1083     -0.41      0.682
## Year2010          0.0948      0.1075      0.88      0.378
## Year2011         -0.0816      0.1261     -0.65      0.518
## Year2012         -0.1100      0.1015     -1.08      0.279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.508
## Multiple R-squared:  0.0389, Adjusted R-squared:  0.0223
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 93 weights are ~= 1. The remaining 908 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0564 0.8600 0.9510 0.8990 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.087 1      1.042
## Year              1.087 16      1.003

```



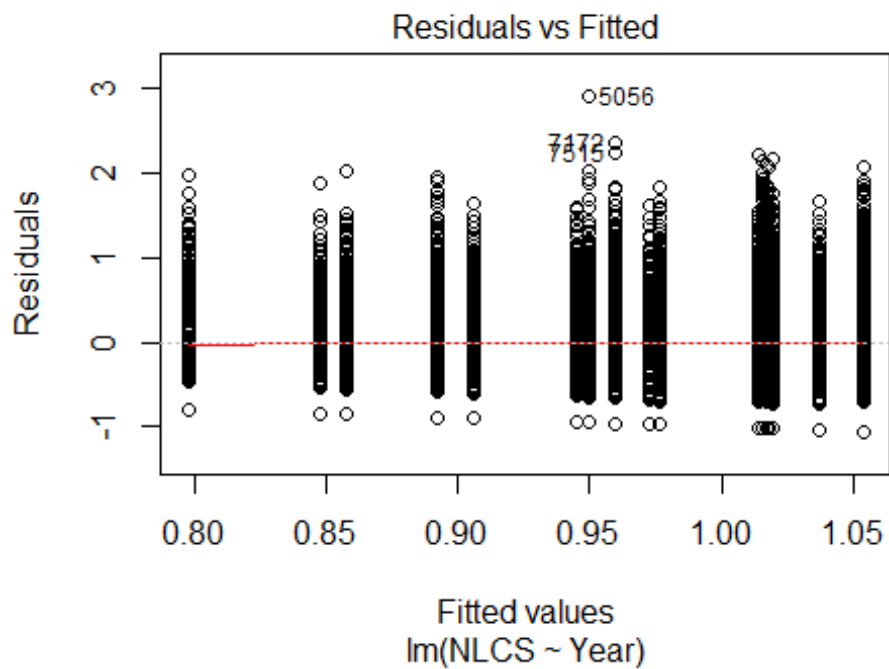
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4265 -0.3283 0.0207 0.3291 2.0702
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.25890 0.08233 15.29 <2e-16 ***
## LastAuthorFemale1 -0.01406 0.04393 -0.32 0.749
## Year1997 -0.16066 0.12847 -1.25 0.211
## Year1998 -0.05759 0.10430 -0.55 0.581
## Year1999 0.04535 0.11760 0.39 0.700
## Year2000 0.13262 0.09609 1.38 0.168
## Year2001 0.22220 0.10473 2.12 0.034 *
## Year2002 0.05899 0.10669 0.55 0.580
## Year2003 0.11870 0.11127 1.07 0.286
## Year2004 -0.00982 0.10137 -0.10 0.923
## Year2005 0.04796 0.11097 0.43 0.666
## Year2006 0.16756 0.10466 1.60 0.110
```

```

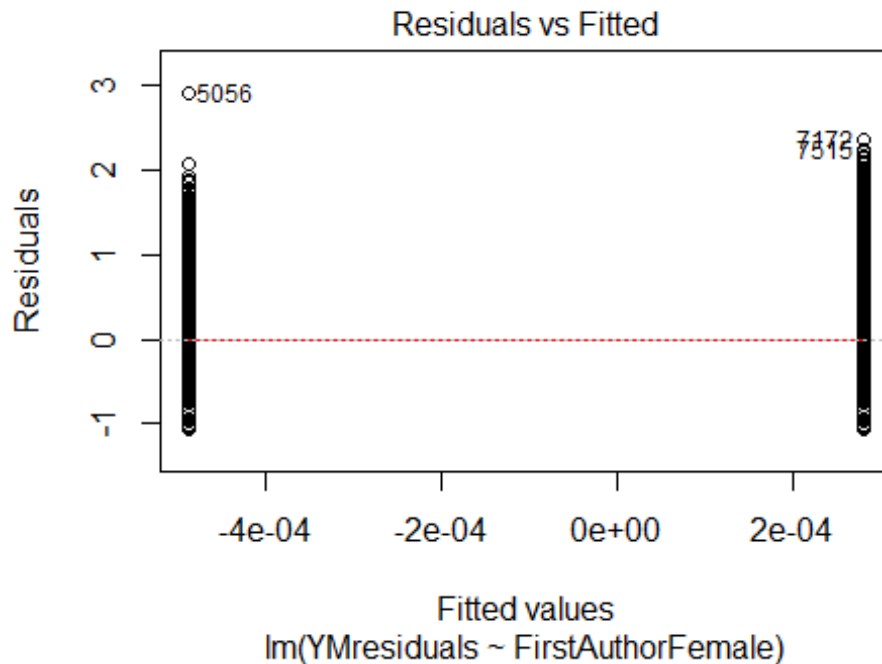
## Year2007      0.09058      0.10666      0.85      0.396
## Year2008      0.09055      0.10074      0.90      0.369
## Year2009     -0.04714      0.10870     -0.43      0.665
## Year2010      0.09273      0.10736      0.86      0.388
## Year2011     -0.07752      0.12644     -0.61      0.540
## Year2012     -0.10992      0.10244     -1.07      0.283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.509
## Multiple R-squared:  0.0355, Adjusted R-squared:  0.0188
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 105 weights are ~= 1. The remaining 896 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0613 0.8610 0.9500 0.8970 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.99e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1001"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3308"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 443 442 479 435 598 530 513 475 460 559 656 866 896 1110 1181
## 2011 2012
## 1312 1366
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 355 367 361 314 412 400 397 394 404 464 566 709 760 942 1028
## 2011 2012

```

```
## 1118 1169
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 345 362 349 306 405 392 386 384 391 453 553 683 732 918 1003
## 2011 2012
## 1091 1137
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 58, df = 16, p-value = 1e-06
```

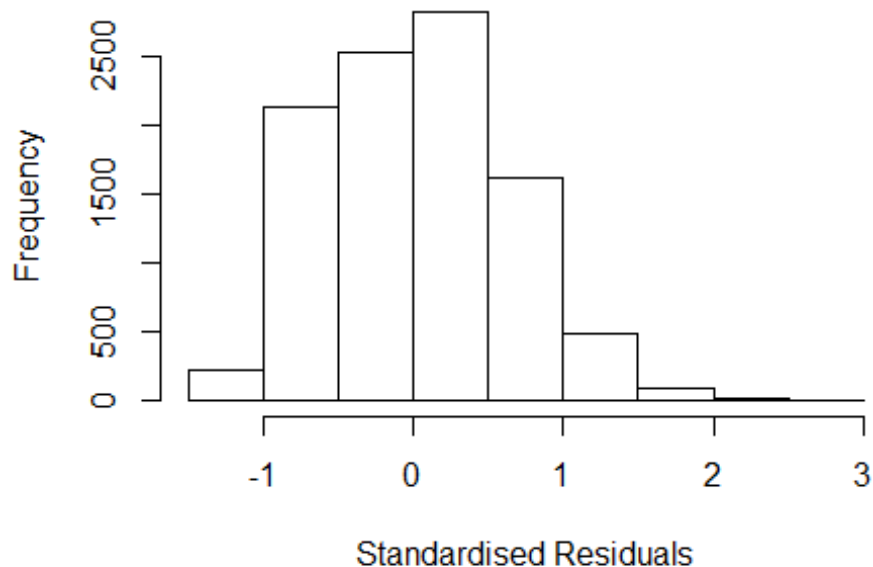


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5.4, df = 1, p-value = 0.02
```



```
## [1] "Female first author team size 2018 geometric mean: 1.45197404400813"
## [1] "Male first author team size 2018 geometric mean: 1.35779845282985"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180000, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.48351010486519"
## [1] "Male last author team size 2018 geometric mean: 1.33276305138206"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190000, p-value = 2e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.515 1          1.586
## LastAuthorFemale  2.503 1          1.582
## UniqueAuthors    1.061 4          1.007
## Year              1.037 16         1.001
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 5056 17544365628 3.848 2005      1200      2      2.966
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3084 -0.4821  0.0146  0.4465  2.9657
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.72330    0.03698   19.56 < 2e-16 ***
## FirstAuthorFemale1 -0.01184    0.02172   -0.55  0.58572
## LastAuthorFemale1  0.00657    0.02182    0.30  0.76345
## UniqueAuthors2    0.17108    0.01841    9.29 < 2e-16 ***
## UniqueAuthors3    0.24937    0.02933    8.50 < 2e-16 ***
## UniqueAuthors4    0.35103    0.04281    8.20 2.7e-16 ***
## UniqueAuthors5    0.27177    0.05157    5.27 1.4e-07 ***
## Year1997         0.09049    0.05231    1.73  0.08367 .
## Year1998         0.05694    0.05101    1.12  0.26435
## Year1999         0.05474    0.04885    1.12  0.26255
```

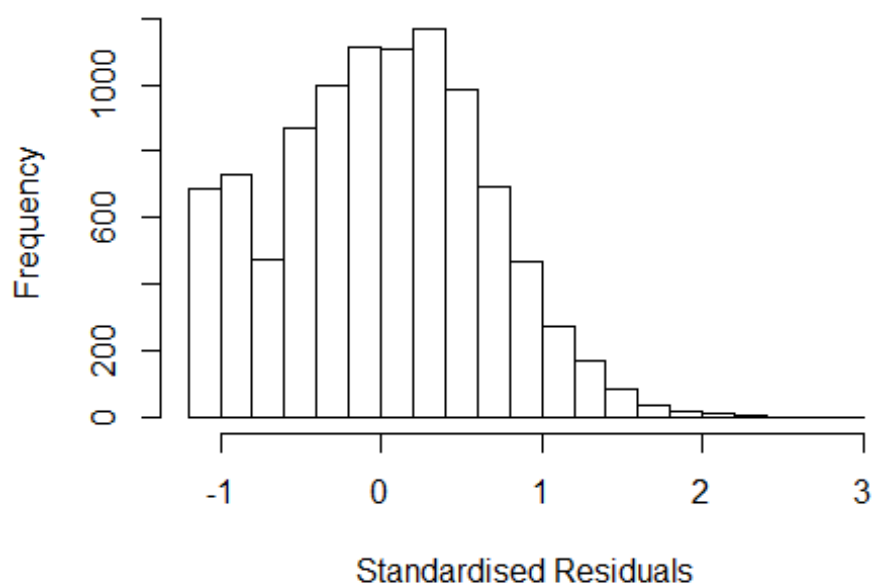


```

## Year2000      0.14299      0.04718      3.03 0.00244 **
## Year2001      0.16386      0.04955      3.31 0.00095 ***
## Year2002      0.22746      0.04929      4.62 4.0e-06 ***
## Year2003      0.19792      0.04698      4.21 2.5e-05 ***
## Year2004      0.25651      0.04832      5.31 1.1e-07 ***
## Year2005      0.16426      0.04632      3.55 0.00039 ***
## Year2006      0.20095      0.04488      4.48 7.7e-06 ***
## Year2007      0.22866      0.04425      5.17 2.4e-07 ***
## Year2008      0.17381      0.04416      3.94 8.3e-05 ***
## Year2009      0.23590      0.04241      5.56 2.7e-08 ***
## Year2010      0.23147      0.04217      5.49 4.1e-08 ***
## Year2011      0.23093      0.04260      5.42 6.1e-08 ***
## Year2012      0.25465      0.04219      6.04 1.6e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.667
## Multiple R-squared:  0.0308, Adjusted R-squared:  0.0286
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 836 weights are ~= 1. The remaining 9054 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0101 0.8680 0.9490 0.9160 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.545 1      1.595
## LastAuthorFemale 2.541 1      1.594
## Year      1.011 16      1.000

```

## Residuals from first and last author



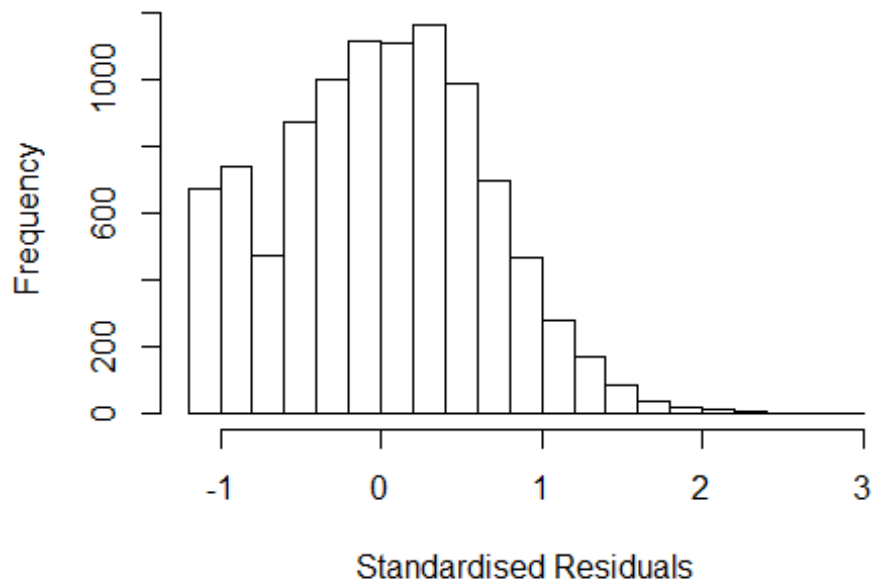
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 5056 17544365628 3.848 2005      1200      2      2.916
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0405 -0.4644  0.0141  0.4554  2.9158
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76433    0.03763   20.31 < 2e-16 ***
## FirstAuthorFemale1 0.00608    0.02208    0.28  0.78309
## LastAuthorFemale1 0.00173    0.02221    0.08  0.93802
## Year1997        0.09746    0.05350    1.82  0.06855 .
## Year1998        0.05980    0.05160    1.16  0.24649
## Year1999        0.06454    0.04951    1.30  0.19247
## Year2000        0.13646    0.04792    2.85  0.00441 **
## Year2001        0.17073    0.05048    3.38  0.00072 ***
## Year2002        0.23726    0.04964    4.78  1.8e-06 ***
## Year2003        0.20455    0.04783    4.28  1.9e-05 ***
## Year2004        0.26402    0.04960    5.32  1.0e-07 ***
## Year2005        0.16008    0.04732    3.38  0.00072 ***
```

```

## Year2006          0.20056    0.04581    4.38  1.2e-05 ***
## Year2007          0.23931    0.04490    5.33  1.0e-07 ***
## Year2008          0.18119    0.04508    4.02  5.9e-05 ***
## Year2009          0.23973    0.04317    5.55  2.9e-08 ***
## Year2010          0.24030    0.04301    5.59  2.4e-08 ***
## Year2011          0.23453    0.04332    5.41  6.3e-08 ***
## Year2012          0.26834    0.04309    6.23  4.9e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.678
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.009
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 800 weights are ~= 1. The remaining 9090 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0245 0.8660 0.9490 0.9170 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.007 1      1.003
## Year              1.007 16      1.000

```

## Residuals from first author



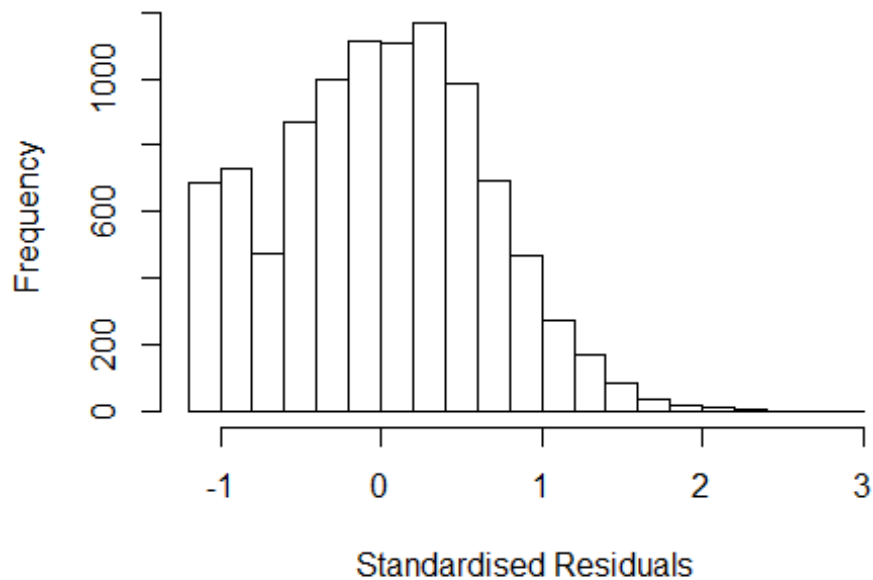
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 5056 17544365628 3.848 2005      1200      2      2.916
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0402 -0.4645  0.0144  0.4552  2.9161
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7644    0.0376   20.32 < 2e-16 ***
## FirstAuthorFemale1 0.0074    0.0139    0.53  0.59427
## Year1997        0.0975    0.0535    1.82  0.06849 .
## Year1998        0.0598    0.0516    1.16  0.24640
## Year1999        0.0645    0.0495    1.30  0.19236
## Year2000        0.1365    0.0479    2.85  0.00441 **
## Year2001        0.1707    0.0505    3.38  0.00072 ***
## Year2002        0.2373    0.0496    4.78  1.8e-06 ***
## Year2003        0.2046    0.0478    4.28  1.9e-05 ***
## Year2004        0.2640    0.0496    5.32  1.0e-07 ***
## Year2005        0.1600    0.0473    3.38  0.00072 ***
## Year2006        0.2006    0.0458    4.38  1.2e-05 ***
```

```

## Year2007          0.2393      0.0449      5.33  1.0e-07 ***
## Year2008          0.1812      0.0451      4.02  5.9e-05 ***
## Year2009          0.2397      0.0432      5.55  2.9e-08 ***
## Year2010          0.2403      0.0430      5.59  2.4e-08 ***
## Year2011          0.2346      0.0433      5.42  6.2e-08 ***
## Year2012          0.2684      0.0431      6.23  4.9e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.678
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.0091
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 805 weights are ~= 1. The remaining 9085 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0244 0.8660 0.9490 0.9170 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.005 1      1.002
## Year      1.005 16      1.000

```

## Residuals from last author



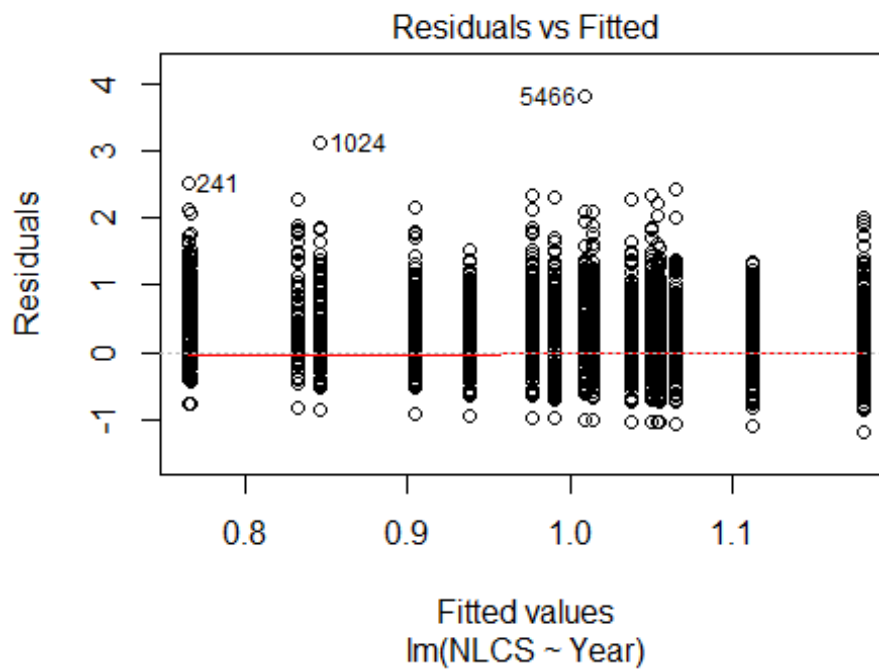
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 5056 17544365628 3.848 2005      1200      2      2.916
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0397 -0.4650  0.0146  0.4551  2.9165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.76484    0.03755  20.37 < 2e-16 ***
## LastAuthorFemale1 0.00641    0.01397   0.46  0.64619
## Year1997        0.09734    0.05350   1.82  0.06888 .
## Year1998        0.05974    0.05160   1.16  0.24701
## Year1999        0.06460    0.04952   1.30  0.19207
## Year2000        0.13651    0.04792   2.85  0.00440 **
## Year2001        0.17077    0.05048   3.38  0.00072 ***
## Year2002        0.23720    0.04964   4.78  1.8e-06 ***
## Year2003        0.20455    0.04783   4.28  1.9e-05 ***
## Year2004        0.26419    0.04958   5.33  1.0e-07 ***
## Year2005        0.16025    0.04732   3.39  0.00071 ***
## Year2006        0.20066    0.04581   4.38  1.2e-05 ***
```

```

## Year2007      0.23942      0.04491      5.33  1.0e-07 ***
## Year2008      0.18120      0.04507      4.02  5.9e-05 ***
## Year2009      0.23994      0.04317      5.56  2.8e-08 ***
## Year2010      0.24024      0.04301      5.59  2.4e-08 ***
## Year2011      0.23453      0.04332      5.41  6.3e-08 ***
## Year2012      0.26844      0.04309      6.23  4.9e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.678
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.00909
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 807 weights are ~= 1. The remaining 9083 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0244 0.8660 0.9490 0.9170 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.01e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 9890"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3309"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 323 302 314 294 335 334 278 262 232 242 324 363 286 318 300
## 2011 2012
## 333 283
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 262 227 232 242 269 263 241 217 193 210 270 316 243 279 251
## 2011 2012

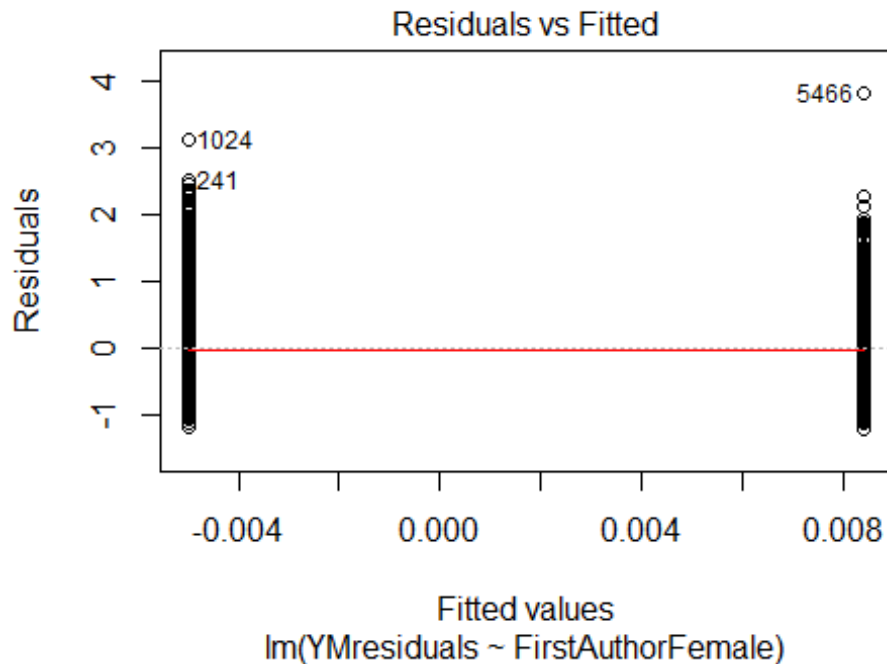
```

```
## 289 235
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 255 220 224 235 261 253 227 202 180 202 256 299 231 264 235
## 2011 2012
## 270 216
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 22, df = 16, p-value = 0.1
```



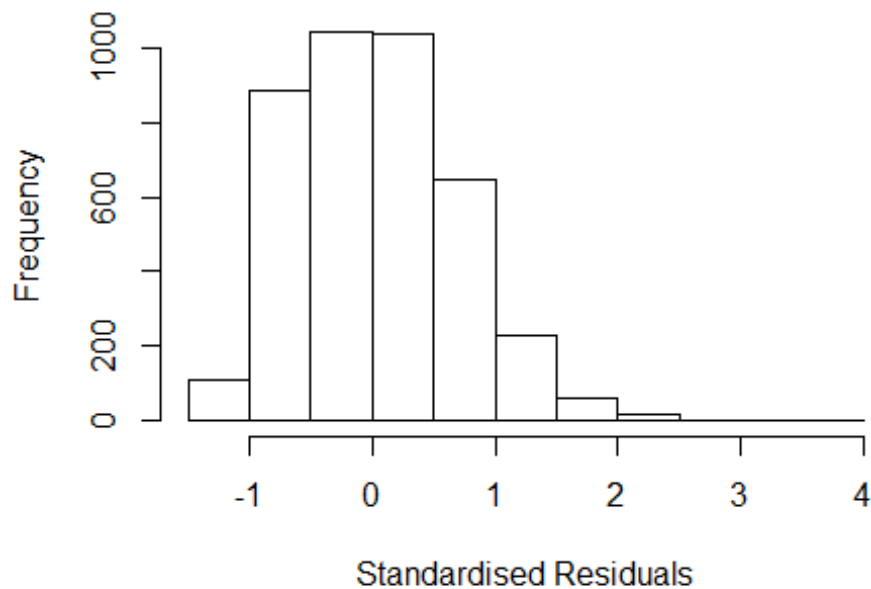
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 11, df = 1, p-value = 7e-04
```





```
## [1] "Female first author team size 2018 geometric mean: 1.77439609117488"
## [1] "Male first author team size 2018 geometric mean: 1.8248784933223"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.83511978482554"
## [1] "Male last author team size 2018 geometric mean: 1.78192314350942"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 4800, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.765 1      1.329
## LastAuthorFemale  1.745 1      1.321
## UniqueAuthors    1.166 4      1.019
## Year              1.171 16     1.005
```

## Residuals from first and last author and team size



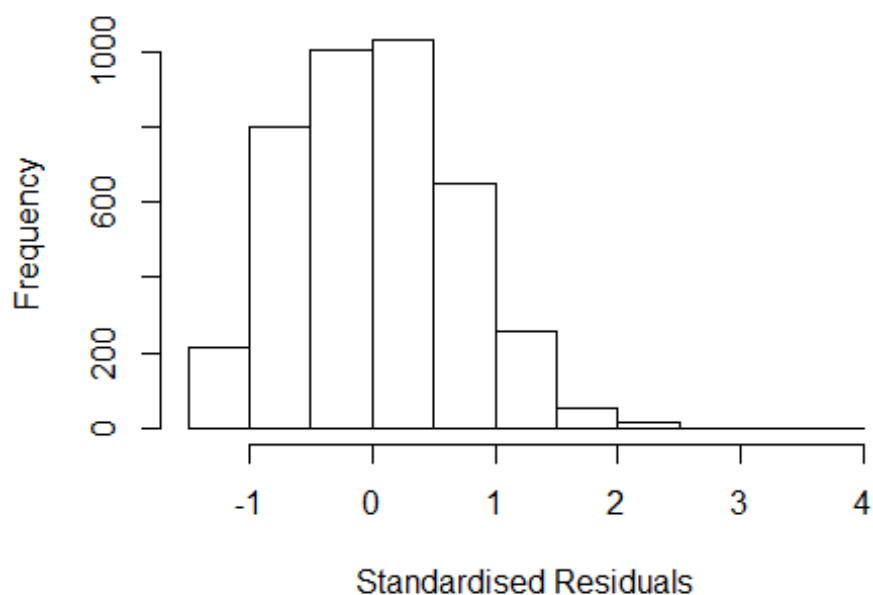
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1024  0033099611 3.971 1999    1706      3      3.299
## 1441  0037667389 3.312 2000    3304      2      2.546
## 2845  3142680761 3.502 2004    1710      3      2.675
## 5466  84861974217 4.826 2012    3309      2      3.658
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.44282 -0.48533 -0.00614  0.46525  3.65813
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.60279    0.04337   13.90 < 2e-16 ***
## FirstAuthorFemale1 -0.01209    0.02859   -0.42  0.67243
## LastAuthorFemale1  0.00467    0.02900    0.16  0.87216
## UniqueAuthors2    0.40252    0.02719   14.80 < 2e-16 ***
## UniqueAuthors3    0.48075    0.03416   14.07 < 2e-16 ***
## UniqueAuthors4    0.46281    0.06235    7.42 1.4e-13 ***
## UniqueAuthors5    0.57378    0.05692   10.08 < 2e-16 ***
```

```

## Year1997      0.06788    0.06065    1.12  0.26309
## Year1998      0.00358    0.05820    0.06  0.95090
## Year1999      0.06886    0.05820    1.18  0.23681
## Year2000      0.16320    0.06049    2.70  0.00701 **
## Year2001      0.10556    0.06284    1.68  0.09306 .
## Year2002      0.18117    0.06457    2.81  0.00504 **
## Year2003      0.27671    0.06250    4.43  9.8e-06 ***
## Year2004      0.22401    0.06682    3.35  0.00081 ***
## Year2005      0.17701    0.06167    2.87  0.00412 **
## Year2006      0.24207    0.05910    4.10  4.3e-05 ***
## Year2007      0.35461    0.05813    6.10  1.2e-09 ***
## Year2008      0.32956    0.06151    5.36  8.9e-08 ***
## Year2009      0.20911    0.06161    3.39  0.00069 ***
## Year2010      0.25758    0.06093    4.23  2.4e-05 ***
## Year2011      0.22862    0.06024    3.79  0.00015 ***
## Year2012      0.16998    0.06232    2.73  0.00640 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.676
## Multiple R-squared:  0.123, Adjusted R-squared:  0.119
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 2 observations c(729,3917) are outliers with |weight| = 0 ( < 2.5e-05);
## 315 weights are ~= 1. The remaining 3713 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.082  0.881   0.945   0.914   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.48e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.790 1          1.338
## LastAuthorFemale  1.789 1          1.337
## Year              1.031 16          1.001

```

## Residuals from first and last author

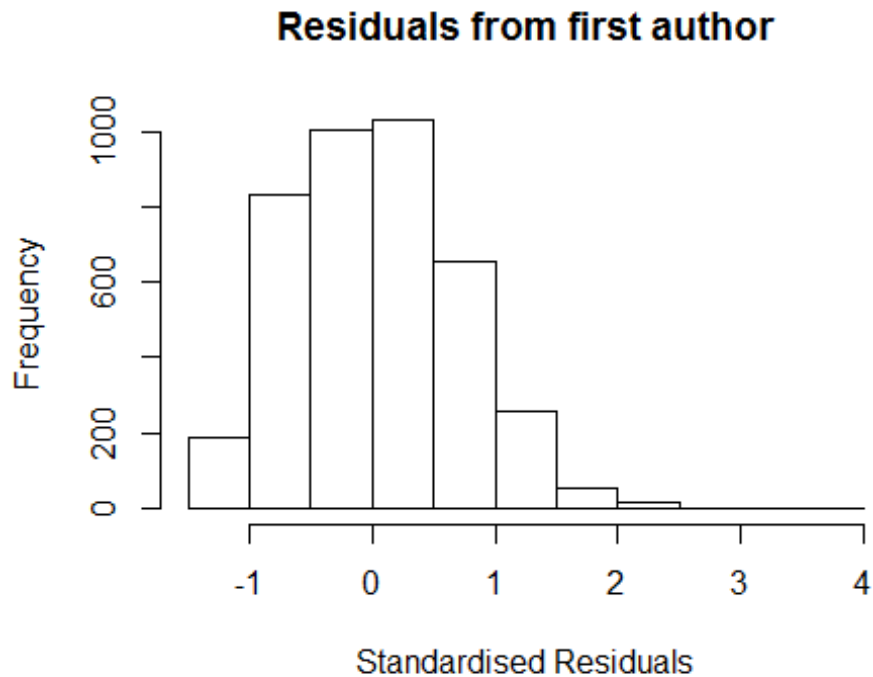


```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 241  21344458965 3.286 1996    1710     2    2.584
## 1024  0033099611 3.971 1999    1706     3    3.178
## 2845  3142680761 3.502 2004    1710     3    2.502
## 5466  84861974217 4.826 2012    3309     2    3.841
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.21329 -0.50548 -0.00177  0.48646  3.84081
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7017    0.0457   15.35 < 2e-16 ***
## FirstAuthorFemale1 0.0524    0.0302    1.74  0.08265 .
## LastAuthorFemale1 -0.0274    0.0309   -0.89  0.37499
## Year1997         0.0709    0.0628    1.13  0.25885
## Year1998         0.0127    0.0611    0.21  0.83501
## Year1999         0.0909    0.0633    1.44  0.15088
## Year2000         0.2197    0.0628    3.50  0.00047 ***
## Year2001         0.1519    0.0651    2.33  0.01974 *
## Year2002         0.2280    0.0678    3.36  0.00078 ***
```

```

## Year2003          0.3562      0.0667      5.34  9.8e-08 ***
## Year2004          0.2979      0.0736      4.05  5.2e-05 ***
## Year2005          0.2223      0.0642      3.46  0.00054 ***
## Year2006          0.3228      0.0625      5.16  2.6e-07 ***
## Year2007          0.4592      0.0606      7.58  4.4e-14 ***
## Year2008          0.4042      0.0649      6.23  5.1e-10 ***
## Year2009          0.2538      0.0638      3.98  7.1e-05 ***
## Year2010          0.3178      0.0637      4.99  6.2e-07 ***
## Year2011          0.3002      0.0647      4.64  3.7e-06 ***
## Year2012          0.2585      0.0683      3.78  0.00016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.721
## Multiple R-squared:  0.0343, Adjusted R-squared:  0.03
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 3917 is an outlier with |weight| = 0 ( < 2.5e-05);
## 322 weights are ~= 1. The remaining 3707 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0129 0.8730 0.9490 0.9170 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.48e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.01 1          1.005
## Year              1.01 16          1.000

```

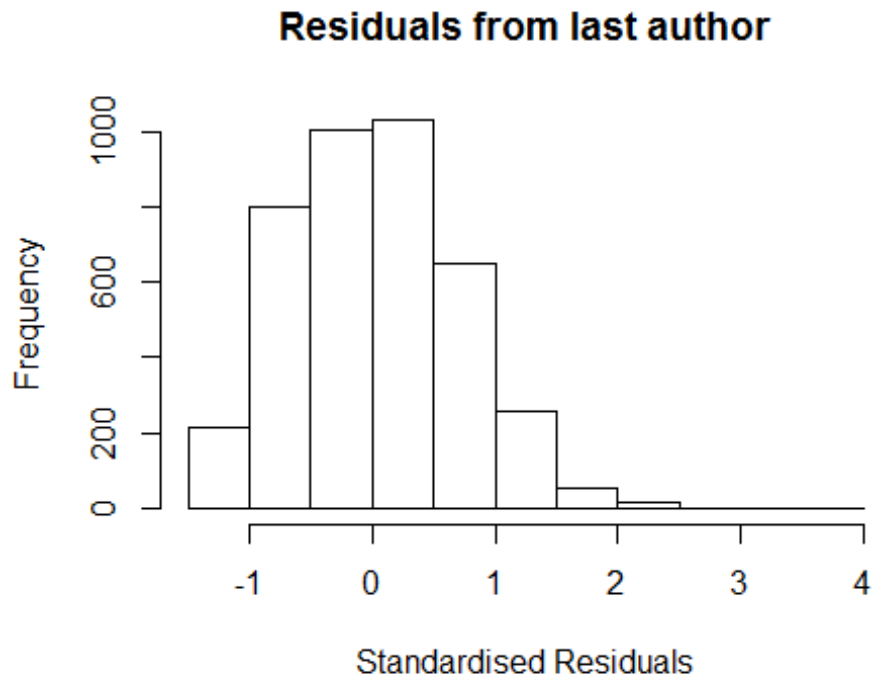


```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 241  21344458965 3.286 1996    1710     2    2.584
## 1024  0033099611 3.971 1999    1706     3    3.178
## 2845  3142680761 3.502 2004    1710     3    2.502
## 5466  84861974217 4.826 2012    3309     2    3.841
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q       Max
## -1.191820 -0.511828 -0.000564  0.487127  3.834745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6991    0.0456   15.33 < 2e-16 ***
## FirstAuthorFemale1 0.0341    0.0227    1.50  0.13406
## Year1997        0.0708    0.0628    1.13  0.25927
## Year1998        0.0136    0.0611    0.22  0.82395
## Year1999        0.0900    0.0633    1.42  0.15539
## Year2000        0.2198    0.0628    3.50  0.00047 ***
## Year2001        0.1528    0.0651    2.35  0.01894 *
## Year2002        0.2295    0.0678    3.39  0.00071 ***
## Year2003        0.3561    0.0667    5.34  1.0e-07 ***
```

```

## Year2004          0.2971      0.0735      4.04  5.4e-05 ***
## Year2005          0.2219      0.0643      3.45  0.00056 ***
## Year2006          0.3230      0.0626      5.16  2.6e-07 ***
## Year2007          0.4587      0.0607      7.56  5.0e-14 ***
## Year2008          0.4042      0.0649      6.23  5.2e-10 ***
## Year2009          0.2548      0.0638      3.99  6.7e-05 ***
## Year2010          0.3177      0.0637      4.99  6.4e-07 ***
## Year2011          0.2982      0.0647      4.61  4.2e-06 ***
## Year2012          0.2581      0.0684      3.78  0.00016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.721
## Multiple R-squared:  0.0342, Adjusted R-squared:  0.0301
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 3917 is an outlier with |weight| = 0 ( < 2.5e-05);
## 333 weights are ~= 1. The remaining 3696 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0129 0.8730 0.9490 0.9160 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.48e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.011 1          1.005
## Year            1.011 16          1.000

```



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 241  21344458965 3.286 1996    1710     2    2.584
## 1024  0033099611 3.971 1999    1706     3    3.178
## 2845  3142680761 3.502 2004    1710     3    2.502
## 5466  84861974217 4.826 2012    3309     2    3.841
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q      Median        3Q        Max
## -1.175784 -0.498595 -0.000909  0.483842  3.850744
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.70759    0.04572   15.48 < 2e-16 ***
## LastAuthorFemale1 0.00873    0.02315    0.38  0.70595
## Year1997        0.07096    0.06278    1.13  0.25844
## Year1998        0.01577    0.06126    0.26  0.79683
## Year1999        0.09105    0.06332    1.44  0.15057
## Year2000        0.22051    0.06277    3.51  0.00045 ***
## Year2001        0.15589    0.06505    2.40  0.01660 *
## Year2002        0.23338    0.06770    3.45  0.00057 ***
## Year2003        0.35651    0.06679    5.34  9.9e-08 ***
```

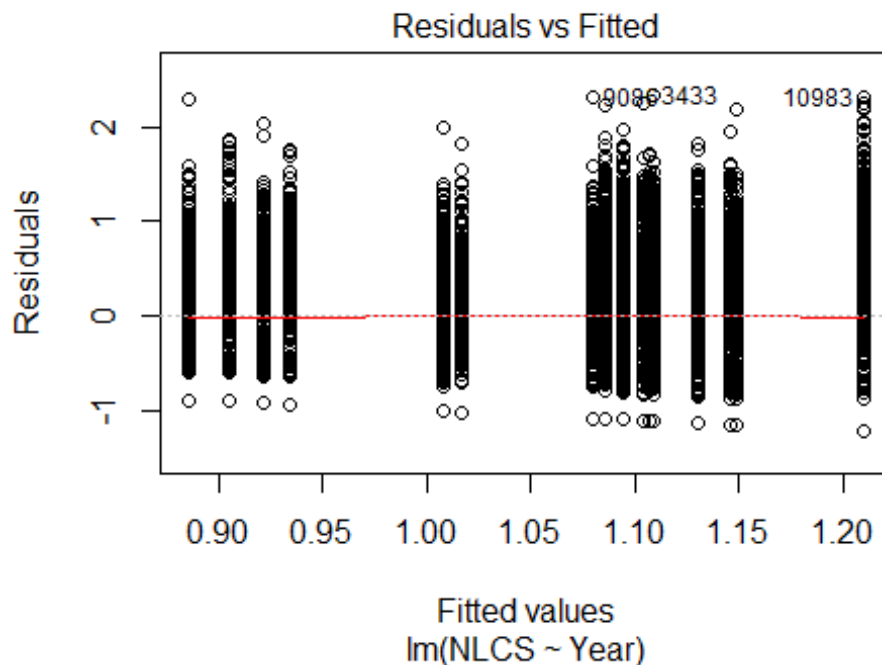


```

## Year2004      0.29839      0.07351      4.06  5.0e-05 ***
## Year2005      0.22371      0.06432      3.48  0.00051 ***
## Year2006      0.32476      0.06273      5.18  2.4e-07 ***
## Year2007      0.45946      0.06070      7.57  4.6e-14 ***
## Year2008      0.40516      0.06495      6.24  4.9e-10 ***
## Year2009      0.25702      0.06387      4.02  5.8e-05 ***
## Year2010      0.31943      0.06381      5.01  5.8e-07 ***
## Year2011      0.29886      0.06478      4.61  4.1e-06 ***
## Year2012      0.25894      0.06846      3.78  0.00016 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.72
## Multiple R-squared:  0.0337, Adjusted R-squared:  0.0296
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 3917 is an outlier with |weight| = 0 ( < 2.5e-05);
## 331 weights are ~ = 1. The remaining 3698 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0136 0.8720 0.9480 0.9160 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.48e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 4030"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3310"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 453 510 471 442 503 469 474 465 573 625 656 750 691 631 644
## 2011 2012
## 755 872

```

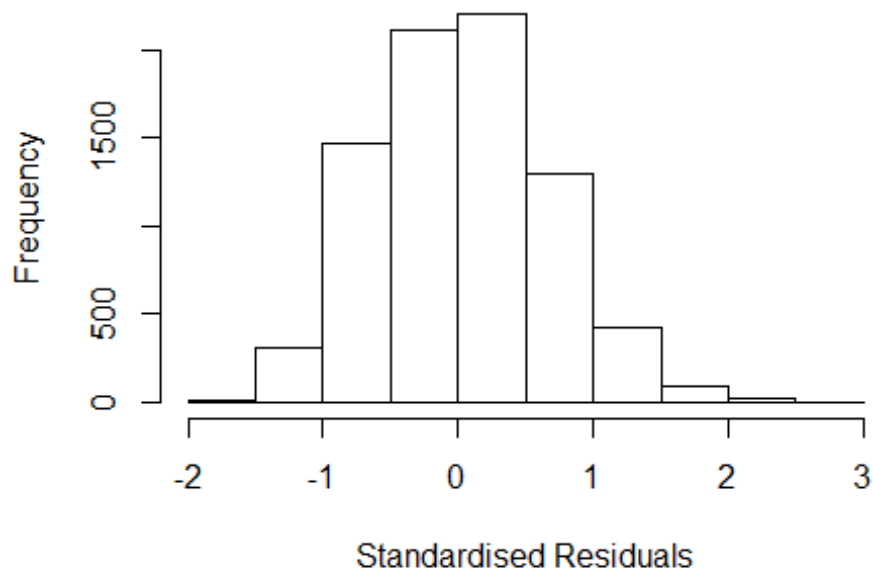
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 312 333 405 367 401 362 410 395 488 525 557 646 593 536 553
## 2011 2012
## 647 743
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 298 326 385 351 394 347 391 379 465 510 541 618 566 518 530
## 2011 2012
## 616 706
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 88, df = 16, p-value = 5e-12
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 13, df = 1, p-value = 4e-04
```



## Residuals from first and last author and team size



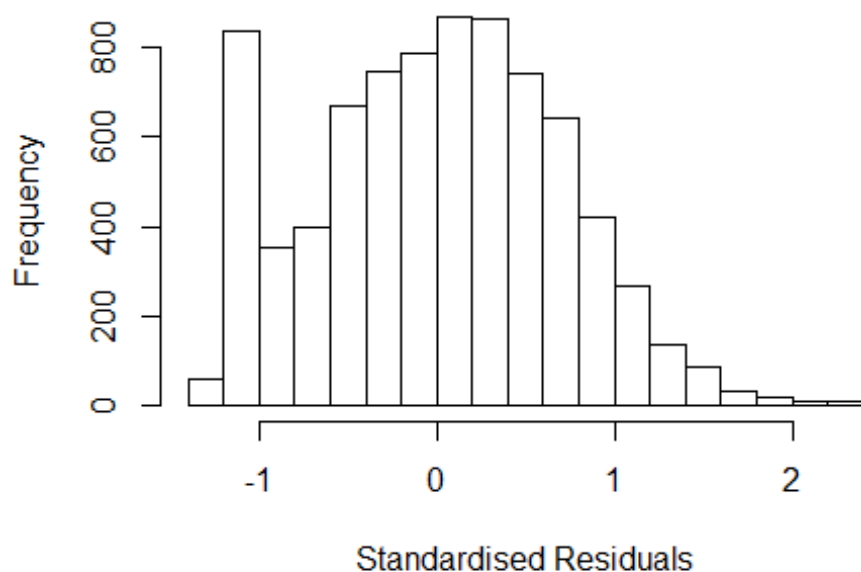
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 3433 0036755089 3.44 2002      1203      3      2.519
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5347 -0.4549  0.0139  0.4591  2.5193
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7375    0.0406   18.17 < 2e-16 ***
## FirstAuthorFemale1 0.0393    0.0200    1.96  0.04975 *
## LastAuthorFemale1 0.0122    0.0201    0.61  0.54434
## UniqueAuthors2    0.4003    0.0198   20.21 < 2e-16 ***
## UniqueAuthors3    0.4753    0.0236   20.12 < 2e-16 ***
## UniqueAuthors4    0.5543    0.0345   16.05 < 2e-16 ***
## UniqueAuthors5    0.6011    0.0471   12.76 < 2e-16 ***
## Year1997          0.0357    0.0561    0.64  0.52417
## Year1998         -0.0311    0.0525   -0.59  0.55421
## Year1999          0.0448    0.0530    0.84  0.39832
```

```

## Year2000          0.1079      0.0512      2.11  0.03498 *
## Year2001          0.1272      0.0519      2.45  0.01422 *
## Year2002          0.1832      0.0530      3.46  0.00055 ***
## Year2003          0.1914      0.0548      3.49  0.00048 ***
## Year2004          0.2313      0.0498      4.64  3.5e-06 ***
## Year2005          0.2432      0.0494      4.92  8.8e-07 ***
## Year2006          0.2282      0.0489      4.67  3.0e-06 ***
## Year2007          0.1877      0.0487      3.86  0.00012 ***
## Year2008          0.1717      0.0491      3.50  0.00047 ***
## Year2009          0.1590      0.0498      3.19  0.00142 **
## Year2010          0.1350      0.0482      2.80  0.00511 **
## Year2011          0.1422      0.0497      2.86  0.00425 **
## Year2012          0.2636      0.0510      5.17  2.4e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.106, Adjusted R-squared:  0.104
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 671 weights are ~= 1. The remaining 7270 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.146  0.860  0.953   0.914  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.039 1      1.428
## LastAuthorFemale  2.035 1      1.427
## Year              1.016 16      1.000

```

## Residuals from first and last author



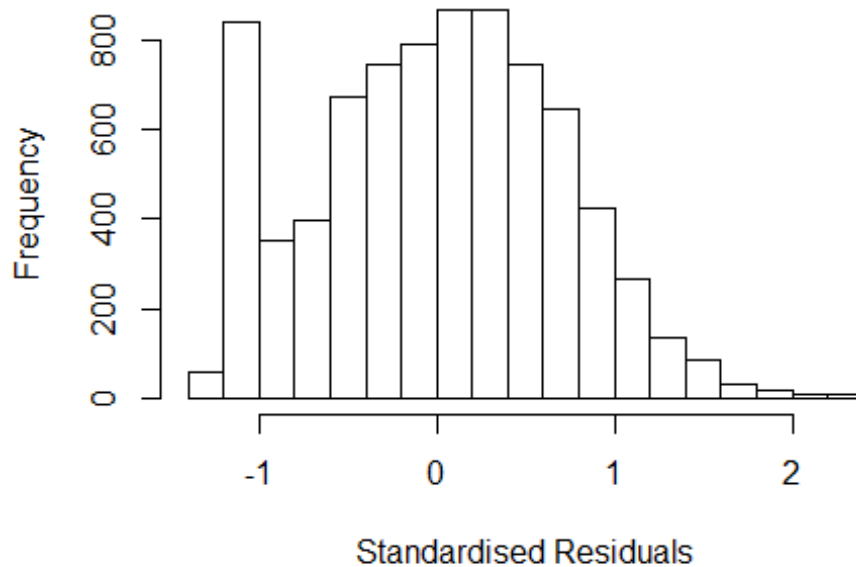
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2211 -0.5018 0.0272 0.4936 2.3890
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8345 0.0421 19.83 < 2e-16 ***
## FirstAuthorFemale1 0.0901 0.0231 3.91 9.5e-05 ***
## LastAuthorFemale1 -0.0032 0.0232 -0.14 0.89028
## Year1997 0.0436 0.0572 0.76 0.44571
## Year1998 -0.0177 0.0546 -0.32 0.74629
## Year1999 0.0409 0.0566 0.72 0.47020
## Year2000 0.1225 0.0532 2.30 0.02139 *
## Year2001 0.1392 0.0533 2.61 0.00906 **
## Year2002 0.2165 0.0558 3.88 0.00010 ***
## Year2003 0.2157 0.0580 3.72 0.00020 ***
## Year2004 0.2469 0.0521 4.74 2.1e-06 ***
## Year2005 0.2736 0.0514 5.32 1.1e-07 ***
```

```

## Year2006          0.2460      0.0506      4.86  1.2e-06 ***
## Year2007          0.2194      0.0508      4.32  1.6e-05 ***
## Year2008          0.1984      0.0511      3.88  0.00010 ***
## Year2009          0.2272      0.0524      4.34  1.5e-05 ***
## Year2010          0.1956      0.0506      3.87  0.00011 ***
## Year2011          0.1843      0.0522      3.53  0.00042 ***
## Year2012          0.2965      0.0537      5.52  3.4e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.727
## Multiple R-squared:  0.0197, Adjusted R-squared:  0.0175
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 678 weights are ~= 1. The remaining 7263 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.259  0.869  0.949  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.008 1          1.004
## Year              1.008 16          1.000

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2184 -0.5023 0.0264 0.4939 2.3894
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8341 0.0420 19.86 < 2e-16 ***
## FirstAuthorFemale1 0.0878 0.0162 5.41 6.5e-08 ***
## Year1997 0.0436 0.0572 0.76 0.44604
## Year1998 -0.0176 0.0546 -0.32 0.74677
## Year1999 0.0409 0.0567 0.72 0.47001
## Year2000 0.1225 0.0532 2.30 0.02137 *
## Year2001 0.1392 0.0533 2.61 0.00908 **
## Year2002 0.2165 0.0558 3.88 0.00011 ***
## Year2003 0.2155 0.0580 3.72 0.00020 ***
## Year2004 0.2468 0.0521 4.74 2.2e-06 ***
## Year2005 0.2735 0.0514 5.32 1.1e-07 ***
## Year2006 0.2460 0.0506 4.86 1.2e-06 ***
```

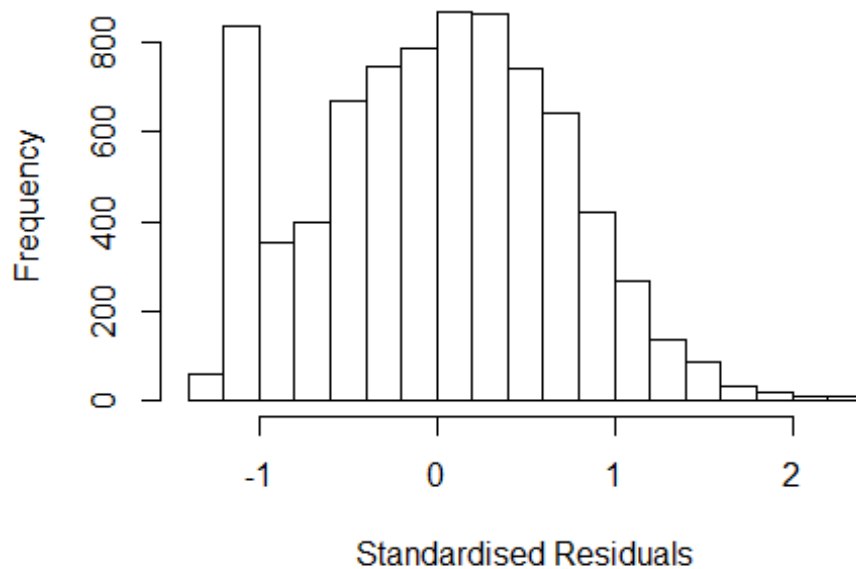


```

## Year2007          0.2195      0.0508      4.32  1.6e-05 ***
## Year2008          0.1983      0.0511      3.88  0.00010 ***
## Year2009          0.2272      0.0524      4.33  1.5e-05 ***
## Year2010          0.1955      0.0506      3.87  0.00011 ***
## Year2011          0.1843      0.0522      3.53  0.00042 ***
## Year2012          0.2965      0.0537      5.52  3.5e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.727
## Multiple R-squared:  0.0197, Adjusted R-squared:  0.0176
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 684 weights are ~= 1. The remaining 7257 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.258  0.869  0.949  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.006 1      1.003
## Year      1.006 16      1.000

```

## Residuals from last author



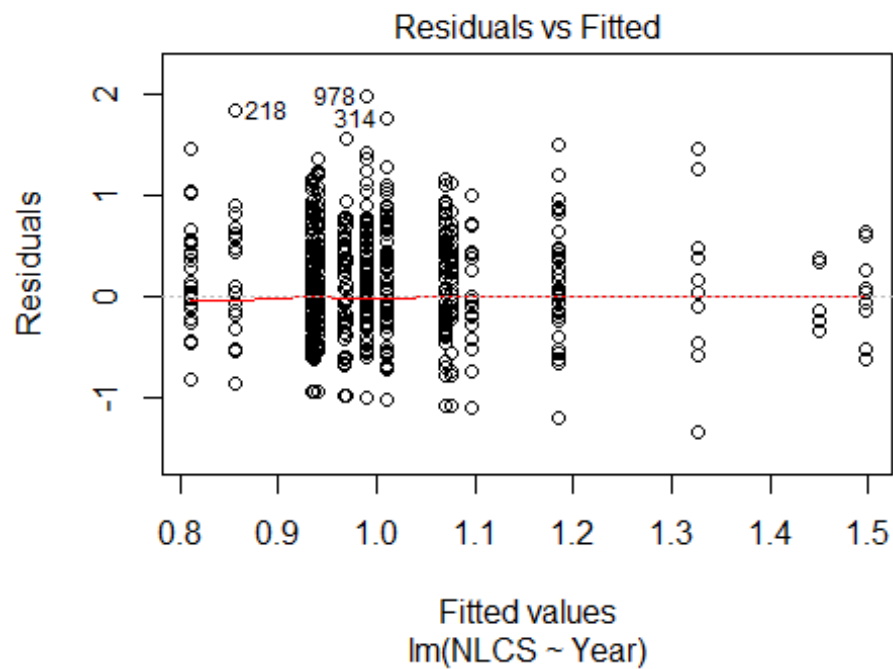
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2060 -0.5009 0.0226 0.4974 2.3768
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8459 0.0422 20.03 < 2e-16 ***
## LastAuthorFemale1 0.0621 0.0163 3.81 0.00014 ***
## Year1997 0.0417 0.0572 0.73 0.46602
## Year1998 -0.0174 0.0547 -0.32 0.75088
## Year1999 0.0388 0.0569 0.68 0.49528
## Year2000 0.1254 0.0535 2.34 0.01908 *
## Year2001 0.1395 0.0533 2.62 0.00890 **
## Year2002 0.2173 0.0560 3.88 0.00010 ***
## Year2003 0.2127 0.0581 3.66 0.00025 ***
## Year2004 0.2467 0.0523 4.72 2.4e-06 ***
## Year2005 0.2737 0.0517 5.30 1.2e-07 ***
## Year2006 0.2475 0.0509 4.86 1.2e-06 ***
```

```

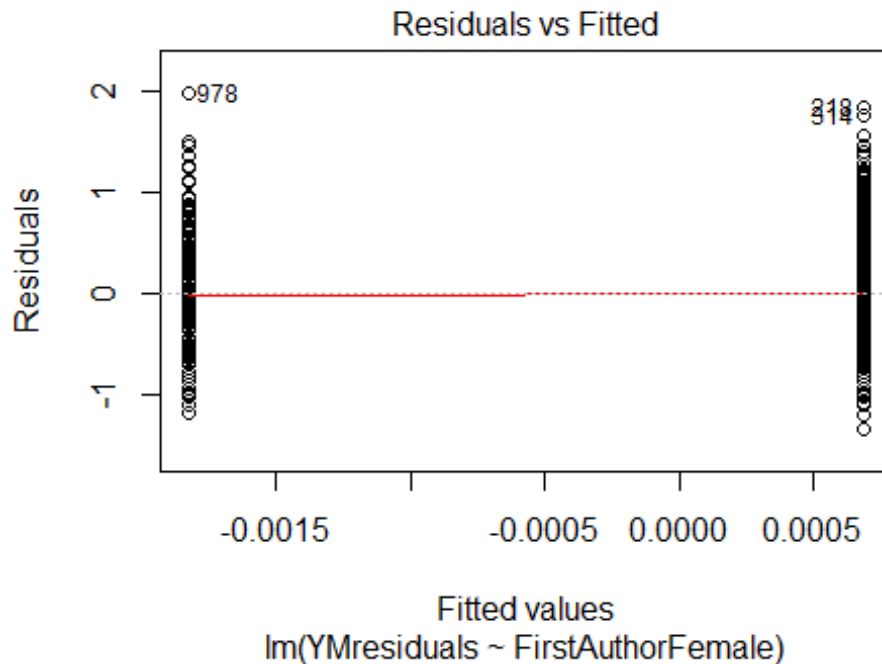
## Year2007          0.2237      0.0511      4.38  1.2e-05 ***
## Year2008          0.1990      0.0513      3.88  0.00011 ***
## Year2009          0.2280      0.0526      4.34  1.5e-05 ***
## Year2010          0.1947      0.0507      3.84  0.00013 ***
## Year2011          0.1875      0.0524      3.58  0.00035 ***
## Year2012          0.2981      0.0539      5.53  3.3e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.727
## Multiple R-squared:  0.0178, Adjusted R-squared:  0.0157
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 664 weights are ~= 1. The remaining 7277 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.264  0.869  0.949  0.917  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.26e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7941"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3311"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   18   20   33   46   48   39   50   50   40   52   52   63   93  121
## 2011 2012
##  146  138
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   11    5   11   25   33   24   26   40   41   29   43   41   46   76  107
## 2011 2012

```

```
## 118 114
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 10 5 11 25 32 22 25 39 40 29 42 39 44 75 102
## 2011 2012
## 114 107
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 17, df = 16, p-value = 0.4
```

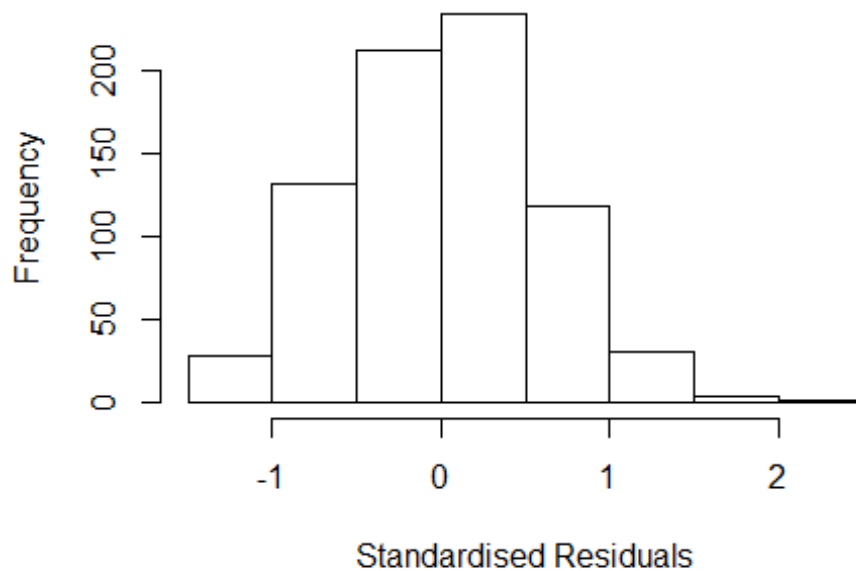


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.019, df = 1, p-value = 0.9
```



```
## [1] "Female first author team size 2018 geometric mean: 1.91530574071499"
## [1] "Male first author team size 2018 geometric mean: 1.9094261894568"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2100, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.95674579215278"
## [1] "Male last author team size 2018 geometric mean: 1.88401536827827"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2300, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.275 1      1.508
## LastAuthorFemale  2.300 1      1.517
## UniqueAuthors    1.517 4      1.053
## Year              1.795 16     1.018
```

## Residuals from first and last author and team size



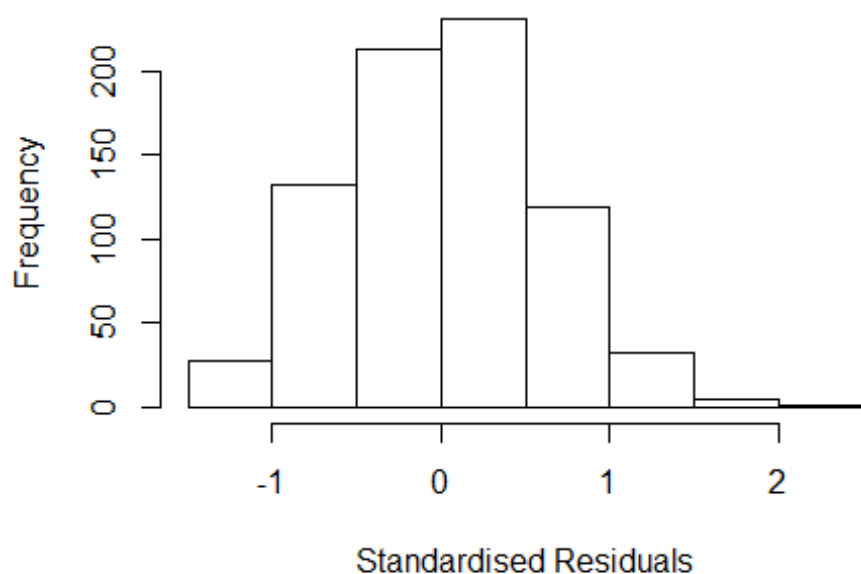
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.445 -0.435 0.013 0.436 2.083
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4669 0.1387 10.57 < 2e-16 ***
## FirstAuthorFemale1 0.0508 0.0747 0.68 0.49701
## LastAuthorFemale1 -0.0969 0.0773 -1.25 0.21028
## UniqueAuthors2 0.0826 0.0664 1.24 0.21422
## UniqueAuthors3 0.1550 0.1225 1.27 0.20610
## UniqueAuthors4 -0.0653 0.1348 -0.48 0.62789
## UniqueAuthors5 0.0806 0.1028 0.78 0.43326
## Year1997 -0.0260 0.1955 -0.13 0.89406
## Year1998 -0.1771 0.3673 -0.48 0.62992
## Year1999 -0.5464 0.1847 -2.96 0.00320 **
```

```

## Year2000          -0.6815      0.1734    -3.93  9.2e-05 ***
## Year2001          -0.4042      0.1843    -2.19  0.02864 *
## Year2002          -0.6491      0.2061    -3.15  0.00170 **
## Year2003          -0.2989      0.1755    -1.70  0.08892 .
## Year2004          -0.4929      0.1898    -2.60  0.00959 **
## Year2005          -0.5447      0.1672    -3.26  0.00117 **
## Year2006          -0.4840      0.1630    -2.97  0.00309 **
## Year2007          -0.4022      0.1701    -2.37  0.01828 *
## Year2008          -0.4725      0.1613    -2.93  0.00350 **
## Year2009          -0.4171      0.1546    -2.70  0.00713 **
## Year2010          -0.5429      0.1513    -3.59  0.00036 ***
## Year2011          -0.5450      0.1487    -3.67  0.00026 ***
## Year2012          -0.5391      0.1520    -3.55  0.00041 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.626
## Multiple R-squared:  0.0434, Adjusted R-squared:  0.0148
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 690 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.246  0.873  0.950  0.913  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.31e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.218 1          1.489
## LastAuthorFemale 2.280 1          1.510
## Year          1.234 16          1.007

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.297 -0.426 0.014 0.422 2.051
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4793 0.1377 10.74 < 2e-16 ***
## FirstAuthorFemale1 0.0647 0.0739 0.87 0.38199
## LastAuthorFemale1 -0.1036 0.0769 -1.35 0.17851
## Year1997 -0.0226 0.1933 -0.12 0.90677
## Year1998 -0.1818 0.3473 -0.52 0.60071
## Year1999 -0.5363 0.1847 -2.90 0.00380 **
## Year2000 -0.6841 0.1743 -3.93 9.5e-05 ***
## Year2001 -0.3943 0.1834 -2.15 0.03192 *
## Year2002 -0.6410 0.2073 -3.09 0.00206 **
## Year2003 -0.2812 0.1761 -1.60 0.11076
## Year2004 -0.4869 0.1888 -2.58 0.01010 *
## Year2005 -0.5488 0.1688 -3.25 0.00120 **
```

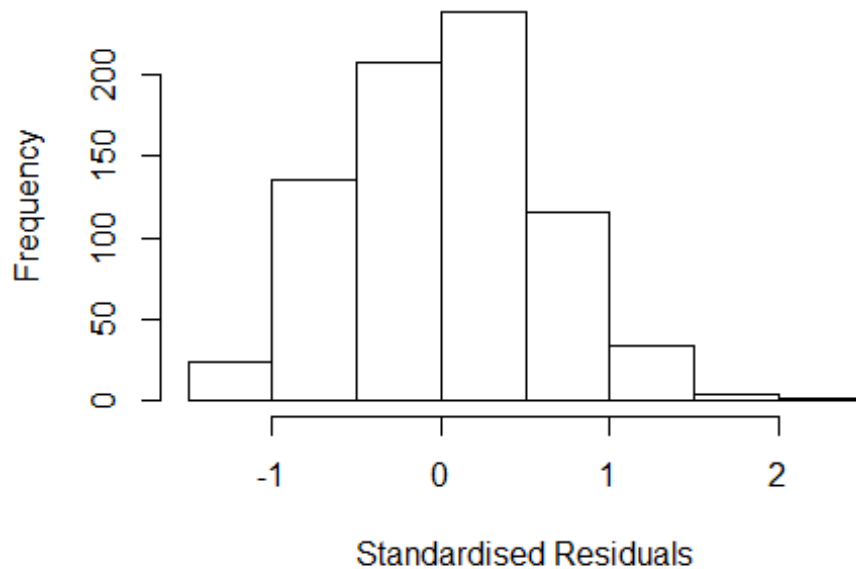


```

## Year2006          -0.4737      0.1649    -2.87   0.00419 **
## Year2007          -0.3947      0.1692    -2.33   0.01994 *
## Year2008          -0.4621      0.1619    -2.85   0.00443 **
## Year2009          -0.4020      0.1555    -2.59   0.00992 **
## Year2010          -0.5434      0.1512    -3.59   0.00035 ***
## Year2011          -0.5372      0.1491    -3.60   0.00034 ***
## Year2012          -0.5263      0.1522    -3.46   0.00058 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.626
## Multiple R-squared:  0.0377, Adjusted R-squared:  0.0143
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 74 weights are ~= 1. The remaining 687 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.261  0.871  0.950  0.912  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.083 1          1.041
## Year              1.083 16          1.003

```

## Residuals from first author



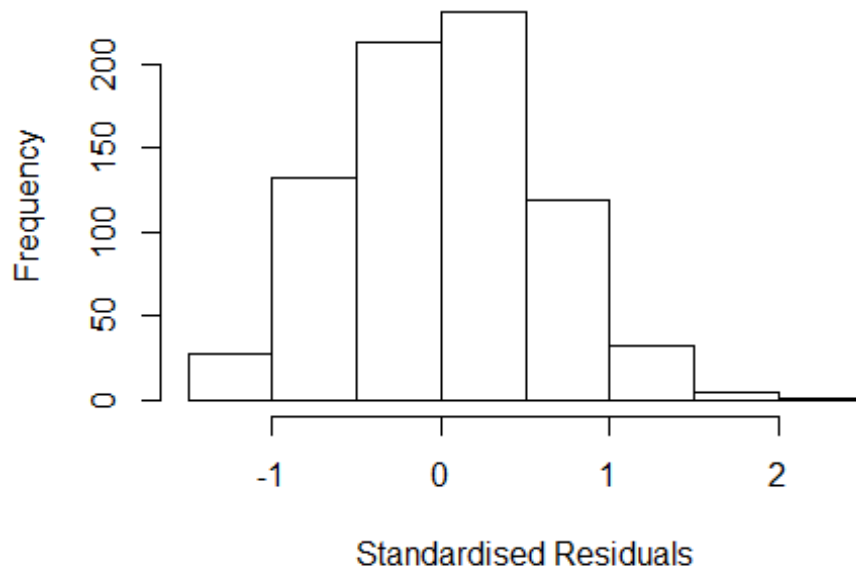
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3204 -0.4372 0.0184 0.4256 2.0273
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.49048 0.14448 10.32 < 2e-16 ***
## FirstAuthorFemale1 -0.00769 0.05169 -0.15 0.88178
## Year1997 -0.04001 0.20035 -0.20 0.84177
## Year1998 -0.17010 0.34229 -0.50 0.61938
## Year1999 -0.55253 0.18957 -2.91 0.00367 **
## Year2000 -0.70229 0.17967 -3.91 0.00010 ***
## Year2001 -0.42000 0.18717 -2.24 0.02513 *
## Year2002 -0.66596 0.21182 -3.14 0.00173 **
## Year2003 -0.30023 0.18189 -1.65 0.09924 .
## Year2004 -0.49561 0.19326 -2.56 0.01053 *
## Year2005 -0.56118 0.17484 -3.21 0.00139 **
## Year2006 -0.50555 0.16888 -2.99 0.00285 **
```

```

## Year2007          -0.40108    0.17508   -2.29  0.02226 *
## Year2008          -0.48109    0.16835   -2.86  0.00439 **
## Year2009          -0.41490    0.16191   -2.56  0.01059 *
## Year2010          -0.56204    0.15701   -3.58  0.00037 ***
## Year2011          -0.55761    0.15514   -3.59  0.00035 ***
## Year2012          -0.54507    0.15788   -3.45  0.00059 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.625
## Multiple R-squared:  0.0356, Adjusted R-squared:  0.0136
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 75 weights are ~= 1. The remaining 686 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.271  0.870   0.948   0.912   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.107 1          1.052
## Year            1.107 16          1.003

```

## Residuals from last author



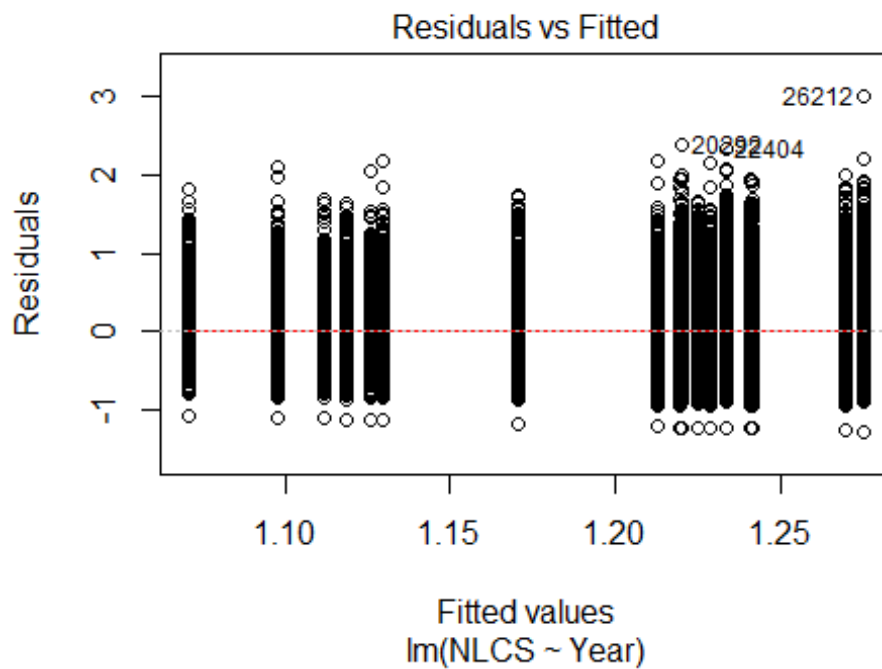
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.316 -0.430 0.012 0.422 2.063
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4933 0.1413 10.57 < 2e-16 ***
## LastAuthorFemale1 -0.0558 0.0537 -1.04 0.29885
## Year1997 -0.0332 0.1948 -0.17 0.86457
## Year1998 -0.1768 0.3466 -0.51 0.61021
## Year1999 -0.5404 0.1875 -2.88 0.00407 **
## Year2000 -0.6942 0.1772 -3.92 9.8e-05 ***
## Year2001 -0.4040 0.1856 -2.18 0.02985 *
## Year2002 -0.6534 0.2104 -3.11 0.00197 **
## Year2003 -0.2934 0.1792 -1.64 0.10198
## Year2004 -0.4876 0.1914 -2.55 0.01106 *
## Year2005 -0.5562 0.1720 -3.23 0.00128 **
## Year2006 -0.4910 0.1671 -2.94 0.00339 **
```

```

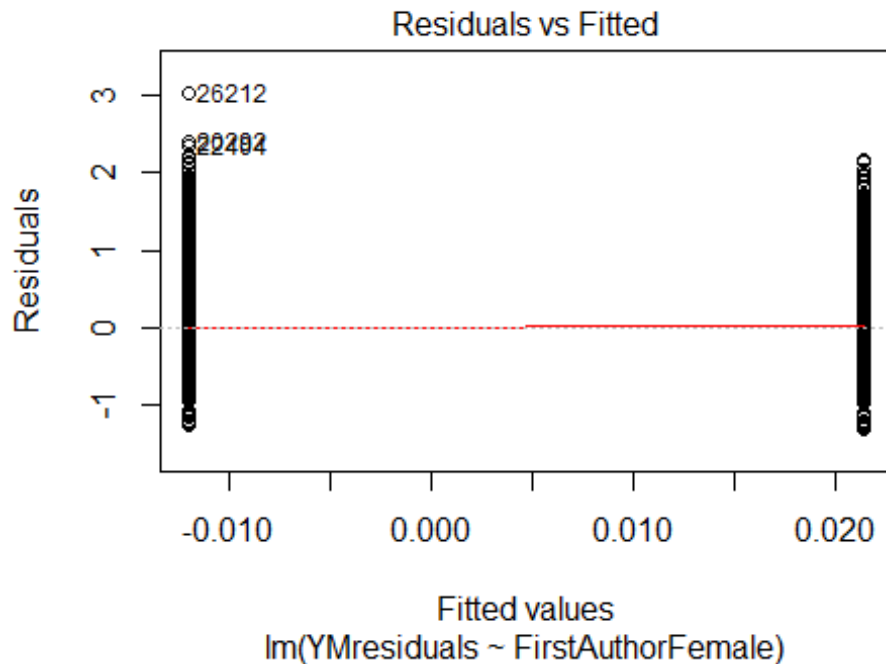
## Year2007          -0.3974      0.1726    -2.30   0.02159 *
## Year2008          -0.4691      0.1659    -2.83   0.00482 **
## Year2009          -0.4074      0.1595    -2.55   0.01082 *
## Year2010          -0.5551      0.1548    -3.58   0.00036 ***
## Year2011          -0.5451      0.1529    -3.57   0.00039 ***
## Year2012          -0.5351      0.1553    -3.45   0.00060 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.627
## Multiple R-squared:  0.0367, Adjusted R-squared:  0.0147
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 71 weights are ~= 1. The remaining 690 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.257  0.870   0.951   0.913   0.985   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.31e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 761"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3312"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 886 860 868 881 1267 1113 1177 1084 1127 1195 1456 1677 1667 2005 2304
## 2011 2012
## 2544 2574
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 776 736 741 735 1010 821 1012 959 989 1040 1254 1459 1452 1761 2017
## 2011 2012

```

```
## 2237 2280
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 759 723 725 722 982 797 989 933 961 1013 1220 1414 1405 1716 1954
## 2011 2012
## 2178 2222
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

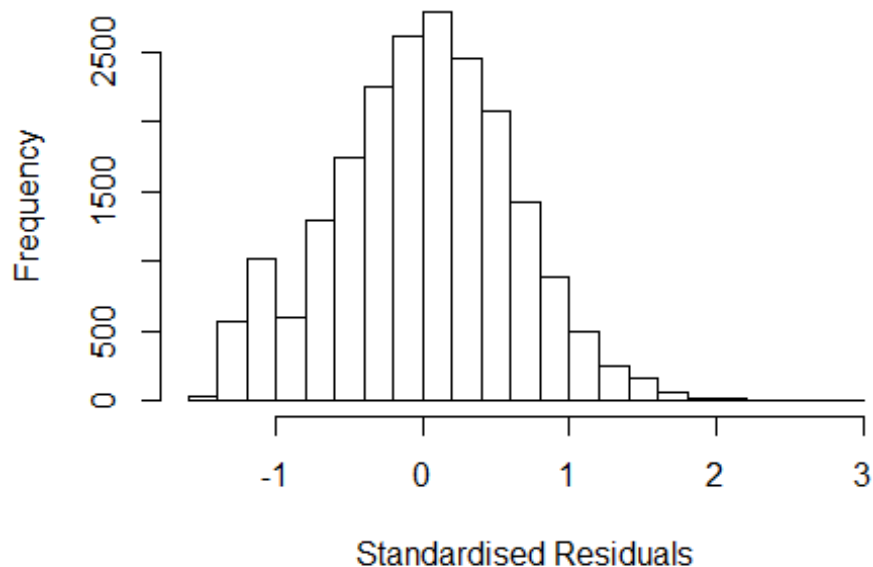


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 17, df = 1, p-value = 4e-05
```



```
## [1] "Female first author team size 2018 geometric mean: 1.46173432147913"
## [1] "Male first author team size 2018 geometric mean: 1.29915032739271"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 720000, p-value = 1e-07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.42477070138044"
## [1] "Male last author team size 2018 geometric mean: 1.3267746444363"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 680000, p-value = 0.005
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.164 1          1.471
## LastAuthorFemale  2.163 1          1.471
## UniqueAuthors    1.029 4          1.004
## Year              1.030 16         1.001
```

## Residuals from first and last author and team size



```
## [1] "List of 1 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 26212 84859894535 4.283 2012      1202      3      2.906
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4740 -0.4061  0.0192  0.4148  2.9057
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.013093   0.023111  43.84 < 2e-16 ***
## FirstAuthorFemale1 0.000392   0.013393   0.03  0.9767
## LastAuthorFemale1 0.038133   0.013462   2.83  0.0046 **
## UniqueAuthors2    0.167669   0.011063  15.16 < 2e-16 ***
## UniqueAuthors3    0.197040   0.017147  11.49 < 2e-16 ***
## UniqueAuthors4    0.200733   0.027768   7.23 5.0e-13 ***
## UniqueAuthors5    0.225837   0.033515   6.74 1.6e-11 ***
## Year1997          0.040623   0.032052   1.27  0.2050
## Year1998          0.059247   0.032068   1.85  0.0647 .
## Year1999          0.009060   0.032409   0.28  0.7798
```

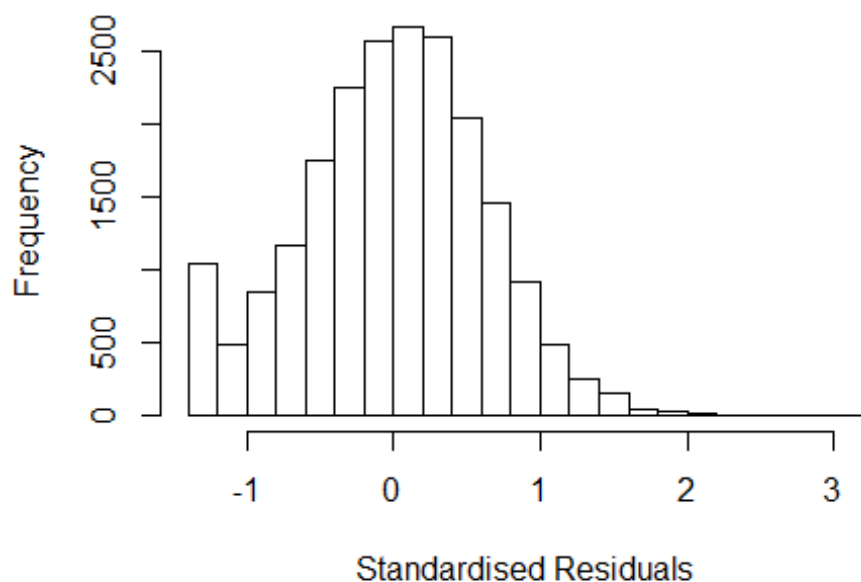


```

## Year2000      0.047308    0.030317    1.56    0.1187
## Year2001      0.043048    0.030946    1.39    0.1642
## Year2002      0.140163    0.030913    4.53    5.8e-06 ***
## Year2003      0.134591    0.030084    4.47    7.7e-06 ***
## Year2004      0.154037    0.030679    5.02    5.2e-07 ***
## Year2005      0.087585    0.030252    2.90    0.0038 **
## Year2006      0.153871    0.028057    5.48    4.2e-08 ***
## Year2007      0.157717    0.027799    5.67    1.4e-08 ***
## Year2008      0.198454    0.028242    7.03    2.2e-12 ***
## Year2009      0.171061    0.027427    6.24    4.6e-10 ***
## Year2010      0.144927    0.027306    5.31    1.1e-07 ***
## Year2011      0.148540    0.027357    5.43    5.7e-08 ***
## Year2012      0.196576    0.027383    7.18    7.3e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.608
## Multiple R-squared:  0.0271, Adjusted R-squared:  0.026
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 20016 is an outlier with |weight| = 0 ( < 4.8e-06);
## 1737 weights are ~= 1. The remaining 18975 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0657 0.8660 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.83e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.231 1      1.494
## LastAuthorFemale 2.235 1      1.495
## Year      1.012 16      1.000

```

## Residuals from first and last author



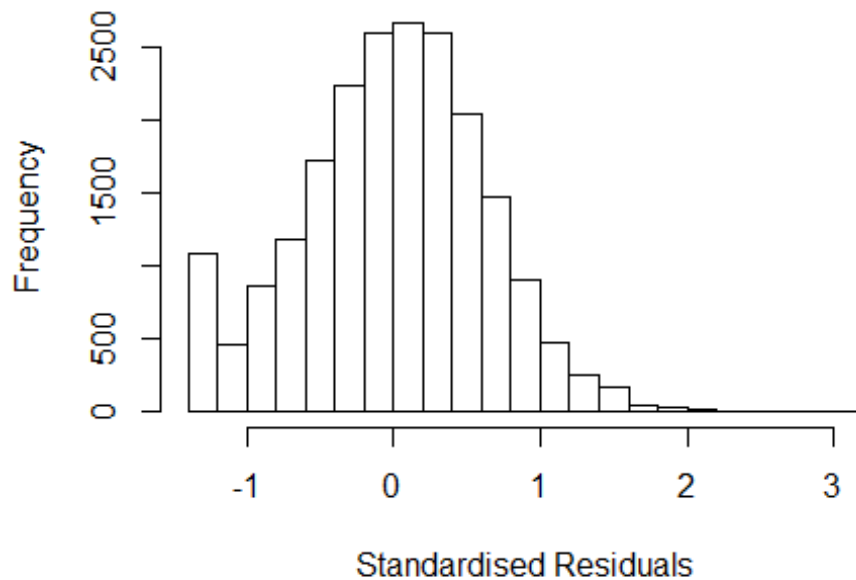
```
## [1] "List of 1 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 26212 84859894535 4.283 2012      1202      3      3.026
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3072 -0.4066  0.0204  0.4155  3.0261
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0533     0.0231  45.62  < 2e-16 ***
## FirstAuthorFemale1  0.0126     0.0137   0.92  0.3568
## LastAuthorFemale1  0.0369     0.0138   2.68  0.0074 **
## Year1997          0.0462     0.0322   1.44  0.1509
## Year1998          0.0610     0.0323   1.89  0.0589 .
## Year1999          0.0243     0.0324   0.75  0.4541
## Year2000          0.0496     0.0305   1.63  0.1037
## Year2001          0.0586     0.0312   1.88  0.0605 .
## Year2002          0.1494     0.0310   4.82 1.4e-06 ***
## Year2003          0.1437     0.0302   4.76 1.9e-06 ***
## Year2004          0.1624     0.0309   5.25 1.5e-07 ***
## Year2005          0.0942     0.0306   3.08  0.0021 **
```

```

## Year2006          0.1585      0.0282      5.63  1.9e-08 ***
## Year2007          0.1701      0.0280      6.09  1.2e-09 ***
## Year2008          0.2042      0.0284      7.19  6.9e-13 ***
## Year2009          0.1749      0.0276      6.35  2.3e-10 ***
## Year2010          0.1503      0.0274      5.48  4.3e-08 ***
## Year2011          0.1560      0.0275      5.67  1.4e-08 ***
## Year2012          0.2035      0.0275      7.39  1.5e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.00957
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 20016 is an outlier with |weight| = 0 ( < 4.8e-06);
## 1761 weights are ~= 1. The remaining 18951 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0884 0.8650 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.83e-06          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.008 1          1.004
## Year          1.008 16          1.000

```

## Residuals from first author



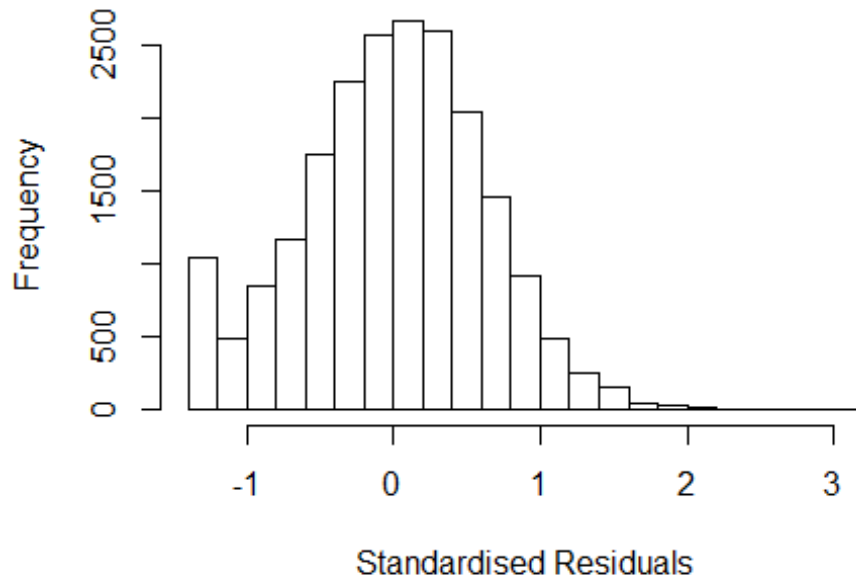
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 26212 84859894535 4.283 2012      1202      3      3.026
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3015 -0.4081  0.0209  0.4154  3.0225
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.05555    0.02307   45.74 < 2e-16 ***
## FirstAuthorFemale1 0.04041    0.00921    4.39 1.2e-05 ***
## Year1997          0.04682    0.03216    1.46  0.1454
## Year1998          0.06052    0.03227    1.88  0.0608 .
## Year1999          0.02499    0.03245    0.77  0.4413
## Year2000          0.05033    0.03048    1.65  0.0987 .
## Year2001          0.05818    0.03122    1.86  0.0624 .
## Year2002          0.15056    0.03098    4.86 1.2e-06 ***
## Year2003          0.14467    0.03020    4.79 1.7e-06 ***
## Year2004          0.16257    0.03094    5.25 1.5e-07 ***
## Year2005          0.09574    0.03063    3.13  0.0018 **
## Year2006          0.15968    0.02818    5.67 1.5e-08 ***
```

```

## Year2007          0.17102      0.02797      6.11  9.9e-10 ***
## Year2008          0.20556      0.02844      7.23  5.1e-13 ***
## Year2009          0.17569      0.02757      6.37  1.9e-10 ***
## Year2010          0.15134      0.02743      5.52  3.5e-08 ***
## Year2011          0.15721      0.02750      5.72  1.1e-08 ***
## Year2012          0.20500      0.02755      7.44  1.0e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0101, Adjusted R-squared:  0.00928
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 20016 is an outlier with |weight| = 0 ( < 4.8e-06);
## 1766 weights are ~= 1. The remaining 18946 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0895 0.8650 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.83e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.01 1      1.005
## Year      1.01 16      1.000

```

## Residuals from last author



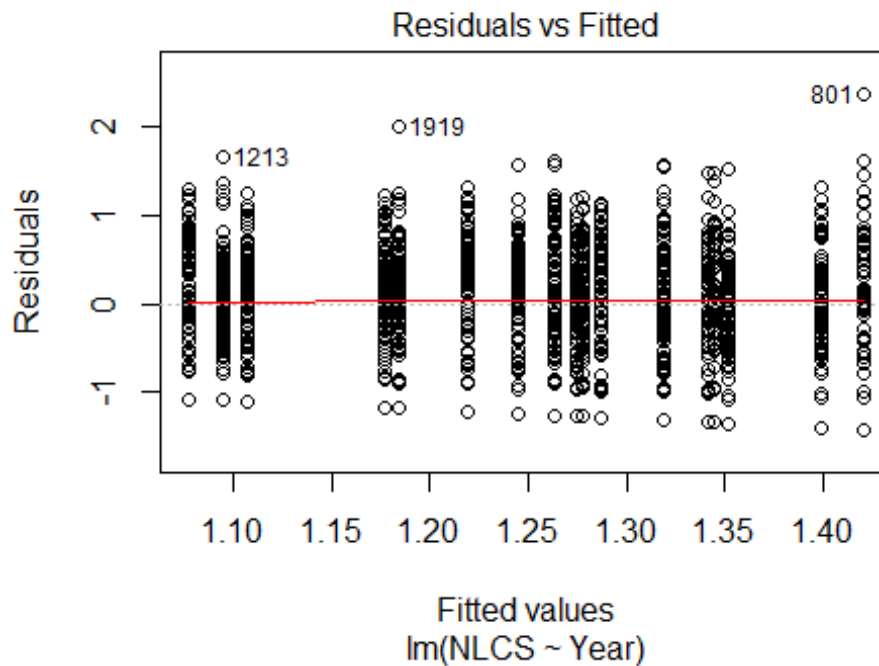
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 26212 84859894535 4.283 2012      1202      3      3.026
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3053 -0.4070  0.0205  0.4150  3.0248
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.05434    0.02307   45.71 < 2e-16 ***
## LastAuthorFemale1 0.04652    0.00927    5.02 5.2e-07 ***
## Year1997        0.04622    0.03216    1.44  0.1507
## Year1998        0.06102    0.03228    1.89  0.0587 .
## Year1999        0.02428    0.03241    0.75  0.4538
## Year2000        0.04979    0.03049    1.63  0.1024
## Year2001        0.05880    0.03120    1.88  0.0595 .
## Year2002        0.14932    0.03096    4.82 1.4e-06 ***
## Year2003        0.14369    0.03019    4.76 2.0e-06 ***
## Year2004        0.16260    0.03091    5.26 1.5e-07 ***
## Year2005        0.09414    0.03061    3.08  0.0021 **
## Year2006        0.15866    0.02816    5.63 1.8e-08 ***
```

```

## Year2007      0.17033      0.02796      6.09  1.1e-09 ***
## Year2008      0.20446      0.02842      7.19  6.5e-13 ***
## Year2009      0.17525      0.02756      6.36  2.1e-10 ***
## Year2010      0.15057      0.02743      5.49  4.1e-08 ***
## Year2011      0.15618      0.02749      5.68  1.4e-08 ***
## Year2012      0.20384      0.02754      7.40  1.4e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.614
## Multiple R-squared:  0.0104, Adjusted R-squared:  0.00958
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 20016 is an outlier with |weight| = 0 ( < 4.8e-06);
## 1771 weights are ~= 1. The remaining 18941 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0889 0.8650 0.9500 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.83e-06      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 20713"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3313"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 129 122 157 152 133 136 141 117 128 124 140 161 126 138 129
## 2011 2012
## 165 194
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 73 72 106 78 75 69 102 93 93 85 92 108 103 105 99

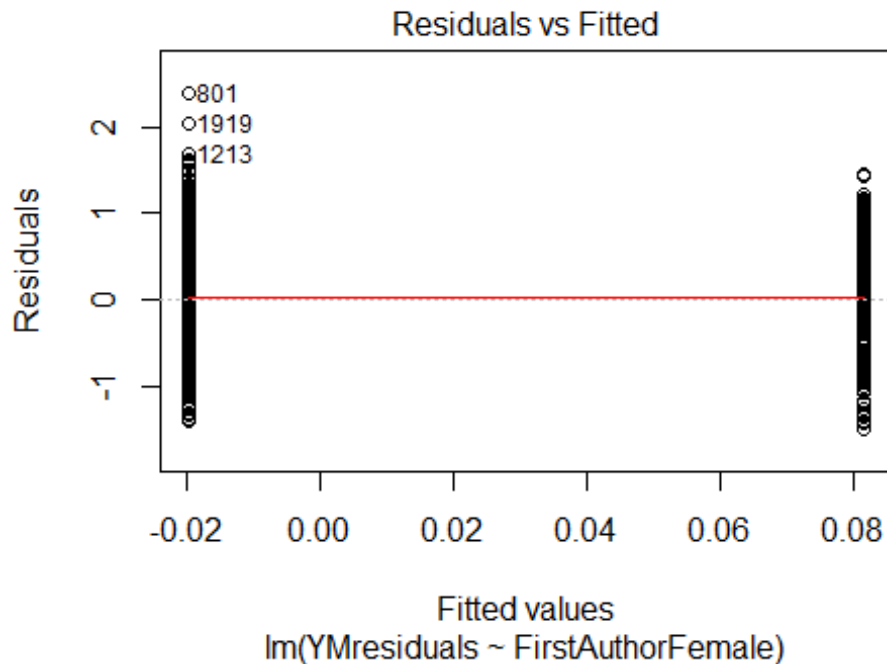
```

```
## 2011 2012
## 129 141
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 71 67 99 77 71 68 100 89 88 80 86 104 92 102 92
## 2011 2012
## 120 132
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 44, df = 16, p-value = 2e-04
```



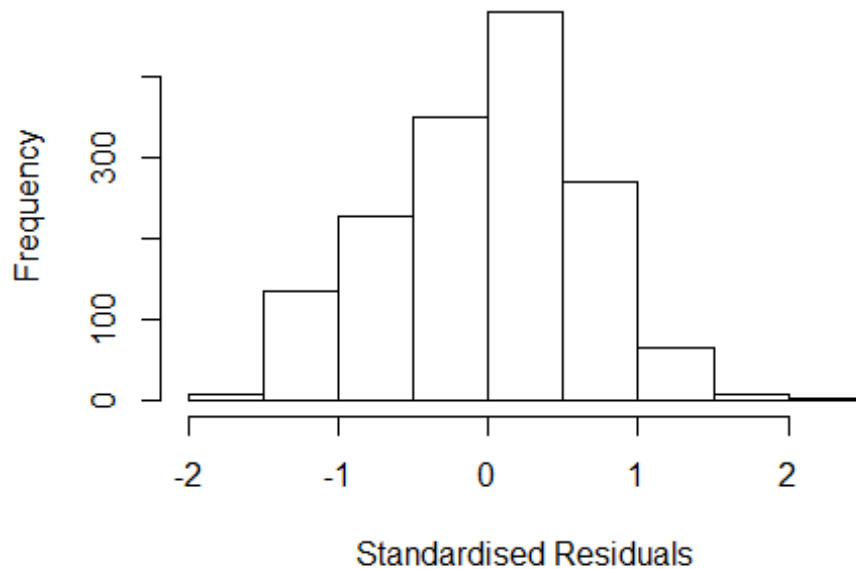
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.4, df = 1, p-value = 0.2
```





```
## [1] "Female first author team size 2018 geometric mean: 2.62001410641333"
## [1] "Male first author team size 2018 geometric mean: 2.29829566720943"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2700, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.46129468265928"
## [1] "Male last author team size 2018 geometric mean: 2.3766211985819"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2200, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.188 1      1.090
## LastAuthorFemale  1.211 1      1.100
## UniqueAuthors     1.242 4      1.027
## Year              1.309 16      1.008
```

## Residuals from first and last author and team size



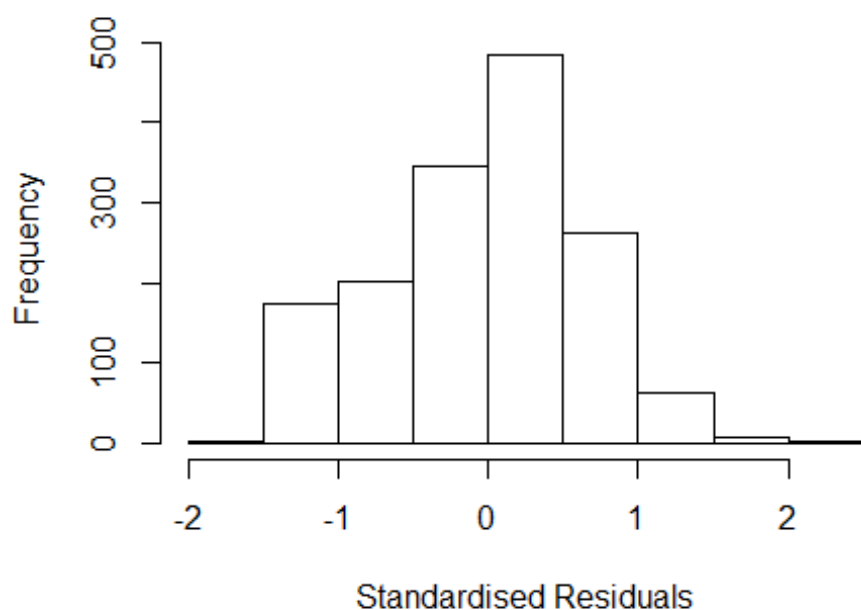
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6350 -0.4635 0.0434 0.4373 2.2690
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2977 0.0919 14.12 < 2e-16 ***
## FirstAuthorFemale1 0.0835 0.0443 1.89 0.05945 .
## LastAuthorFemale1 0.0414 0.0469 0.88 0.37717
## UniqueAuthors2 0.3373 0.0409 8.24 3.7e-16 ***
## UniqueAuthors3 0.3563 0.0499 7.14 1.4e-12 ***
## UniqueAuthors4 0.3629 0.0642 5.65 1.9e-08 ***
## UniqueAuthors5 0.2697 0.1116 2.42 0.01577 *
## Year1997 -0.0983 0.1347 -0.73 0.46581
## Year1998 -0.1919 0.1286 -1.49 0.13601
## Year1999 -0.1165 0.1243 -0.94 0.34848
```

```

## Year2000          -0.1544      0.1151   -1.34   0.17985
## Year2001          -0.0426      0.1410   -0.30   0.76257
## Year2002          -0.2153      0.1189   -1.81   0.07034 .
## Year2003          -0.3989      0.1199   -3.33   0.00090 ***
## Year2004          -0.3909      0.1104   -3.54   0.00041 ***
## Year2005          -0.1315      0.1045   -1.26   0.20858
## Year2006          -0.2254      0.1112   -2.03   0.04282 *
## Year2007          -0.2345      0.1125   -2.08   0.03738 *
## Year2008          -0.3747      0.1121   -3.34   0.00085 ***
## Year2009          -0.3359      0.1123   -2.99   0.00284 **
## Year2010          -0.2659      0.1085   -2.45   0.01439 *
## Year2011          -0.3475      0.1115   -3.12   0.00186 **
## Year2012          -0.1874      0.1072   -1.75   0.08063 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.651
## Multiple R-squared:  0.0864, Adjusted R-squared:  0.0731
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 141 weights are ~= 1. The remaining 1397 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.200  0.860  0.946  0.906  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          6.50e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.228 1          1.108
## LastAuthorFemale  1.211 1          1.101
## Year              1.114 16          1.003

```

## Residuals from first and last author

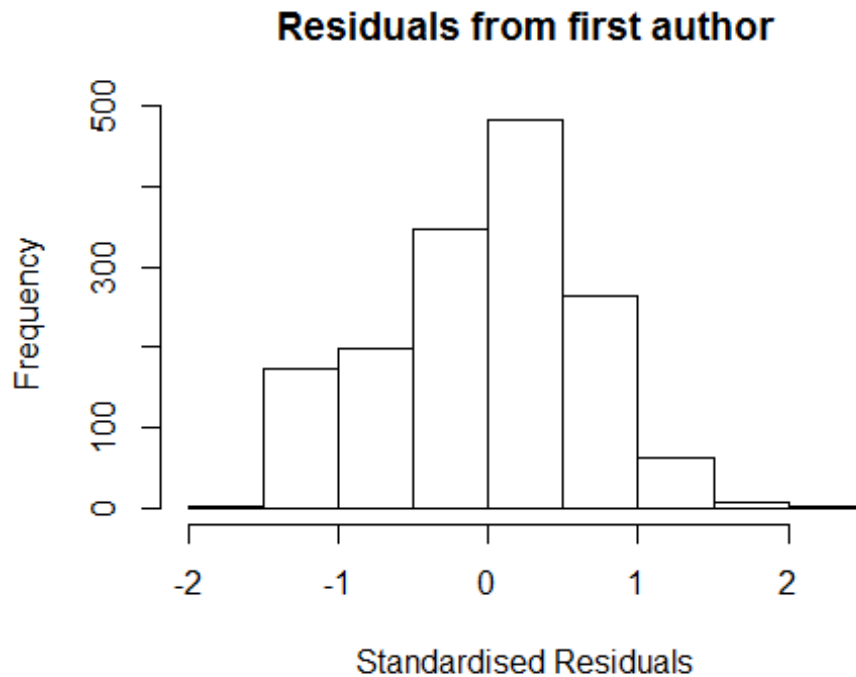


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5524 -0.4715  0.0688  0.4518  2.3868
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.4154     0.0911   15.53  <2e-16 ***
## FirstAuthorFemale1  0.0912     0.0472    1.93   0.0535 .
## LastAuthorFemale1  0.0458     0.0494    0.93   0.3542
## Year1997          -0.0284     0.1354   -0.21   0.8337
## Year1998          -0.2003     0.1327   -1.51   0.1313
## Year1999          -0.1235     0.1266   -0.98   0.3292
## Year2000          -0.0968     0.1166   -0.83   0.4066
## Year2001          -0.0182     0.1417   -0.13   0.8979
## Year2002          -0.1739     0.1213   -1.43   0.1519
## Year2003          -0.3453     0.1228   -2.81   0.0050 **
## Year2004          -0.3635     0.1141   -3.19   0.0015 **
## Year2005          -0.0734     0.1065   -0.69   0.4907
```

```

## Year2006          -0.1690      0.1128   -1.50   0.1341
## Year2007          -0.1589      0.1142   -1.39   0.1642
## Year2008          -0.2863      0.1131   -2.53   0.0115 *
## Year2009          -0.2999      0.1145   -2.62   0.0089 **
## Year2010          -0.1696      0.1108   -1.53   0.1262
## Year2011          -0.2682      0.1125   -2.38   0.0172 *
## Year2012          -0.1152      0.1073   -1.07   0.2831
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.0274, Adjusted R-squared:  0.0159
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 117 weights are ~= 1. The remaining 1421 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.163  0.863  0.949  0.905  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1      1.029
## Year              1.058 16      1.002

```

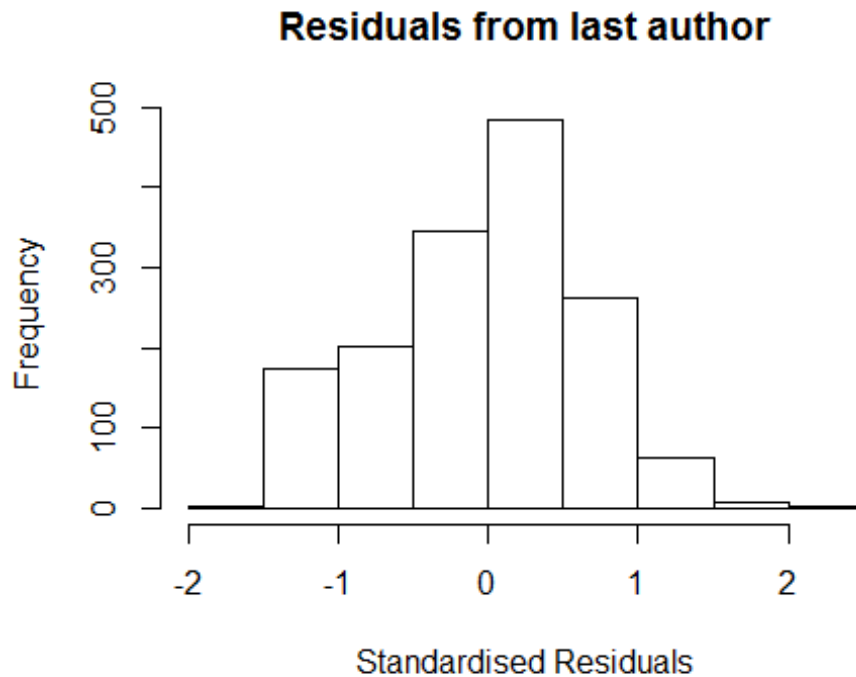


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.532 -0.473 0.070 0.450 2.384
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4180 0.0907 15.64 <2e-16 ***
## FirstAuthorFemale1 0.1135 0.0442 2.57 0.0103 *
## Year1997 -0.0273 0.1352 -0.20 0.8403
## Year1998 -0.1994 0.1321 -1.51 0.1314
## Year1999 -0.1207 0.1263 -0.96 0.3394
## Year2000 -0.0920 0.1161 -0.79 0.4284
## Year2001 -0.0182 0.1416 -0.13 0.8975
## Year2002 -0.1716 0.1209 -1.42 0.1562
## Year2003 -0.3451 0.1223 -2.82 0.0048 **
## Year2004 -0.3630 0.1137 -3.19 0.0014 **
## Year2005 -0.0679 0.1058 -0.64 0.5213
## Year2006 -0.1641 0.1122 -1.46 0.1438
```

```

## Year2007          -0.1599      0.1140    -1.40    0.1609
## Year2008          -0.2847      0.1128    -2.52    0.0117 *
## Year2009          -0.2982      0.1141    -2.61    0.0091 **
## Year2010          -0.1671      0.1103    -1.51    0.1300
## Year2011          -0.2647      0.1119    -2.37    0.0181 *
## Year2012          -0.1137      0.1069    -1.06    0.2876
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.661
## Multiple R-squared:  0.0268, Adjusted R-squared:  0.0159
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 114 weights are ~= 1. The remaining 1424 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.166  0.864  0.950  0.905  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.046 1          1.023
## Year              1.046 16          1.001

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5147 -0.4730 0.0628 0.4482 2.3785
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4245 0.0918 15.52 <2e-16 ***
## LastAuthorFemale1 0.0902 0.0457 1.97 0.0486 *
## Year1997 -0.0296 0.1362 -0.22 0.8278
## Year1998 -0.2034 0.1332 -1.53 0.1270
## Year1999 -0.1172 0.1266 -0.93 0.3549
## Year2000 -0.1040 0.1173 -0.89 0.3754
## Year2001 -0.0189 0.1423 -0.13 0.8942
## Year2002 -0.1737 0.1219 -1.42 0.1545
## Year2003 -0.3530 0.1239 -2.85 0.0044 **
## Year2004 -0.3570 0.1143 -3.12 0.0018 **
## Year2005 -0.0779 0.1071 -0.73 0.4673
## Year2006 -0.1708 0.1132 -1.51 0.1316
```

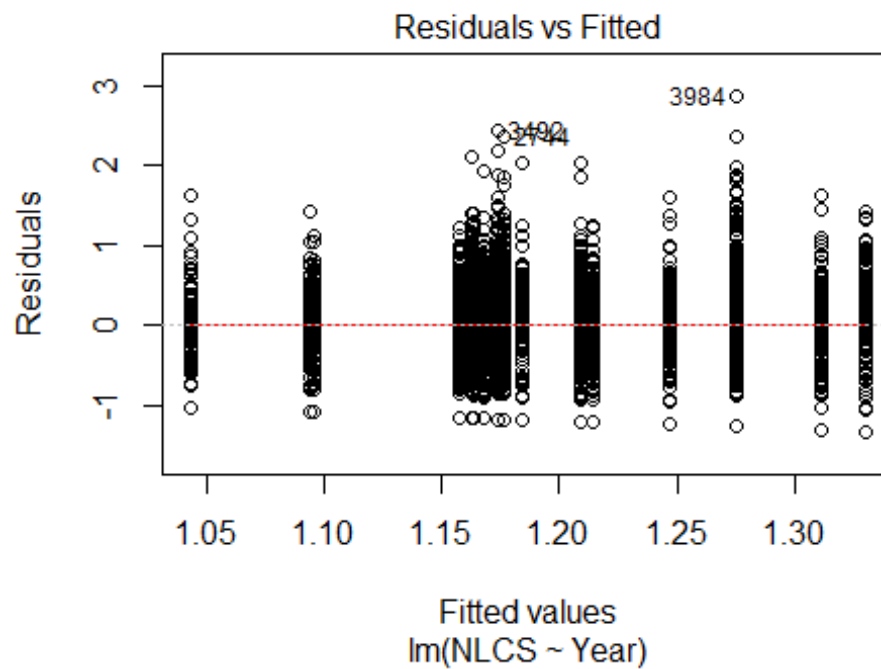


```

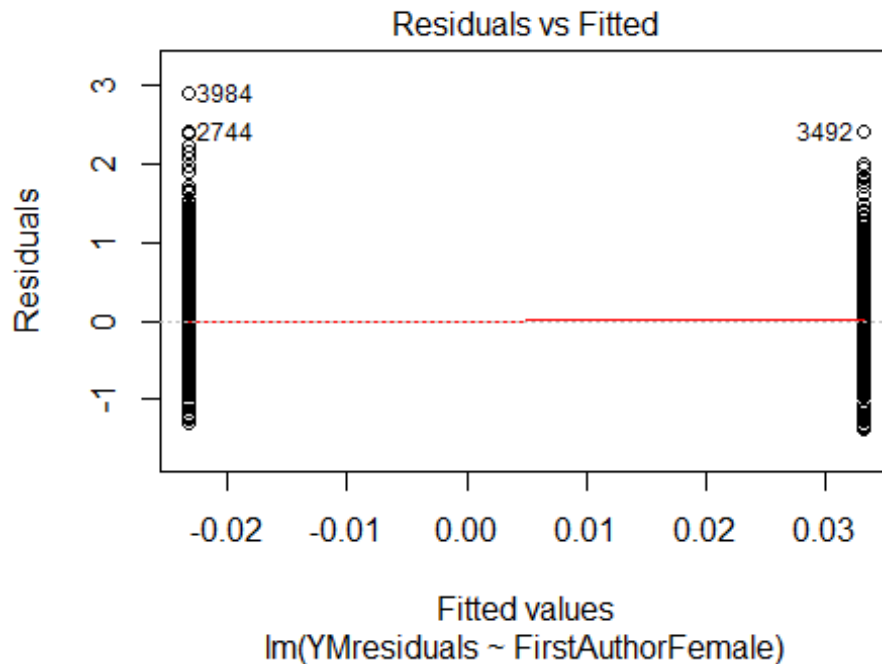
## Year2007          -0.1585      0.1149   -1.38    0.1679
## Year2008          -0.2847      0.1136   -2.51    0.0123 *
## Year2009          -0.3032      0.1153   -2.63    0.0086 **
## Year2010          -0.1670      0.1115   -1.50    0.1344
## Year2011          -0.2605      0.1126   -2.31    0.0208 *
## Year2012          -0.1106      0.1076   -1.03    0.3040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.66
## Multiple R-squared:  0.0254, Adjusted R-squared:  0.0145
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 126 weights are ~ = 1. The remaining 1412 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.166  0.859  0.947   0.904  0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.50e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1538"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3314"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 128  92  121  123  138  178  163  134  173  170  193  245  261  274  391
## 2011 2012
## 428  504
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 98  71  97  98  105  120  132  111  147  144  165  202  218  224  324
## 2011 2012

```

```
## 357 418
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 94 69 93 95 104 118 130 105 144 141 159 194 212 210 313
## 2011 2012
## 344 408
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 120, df = 16, p-value <2e-16
```

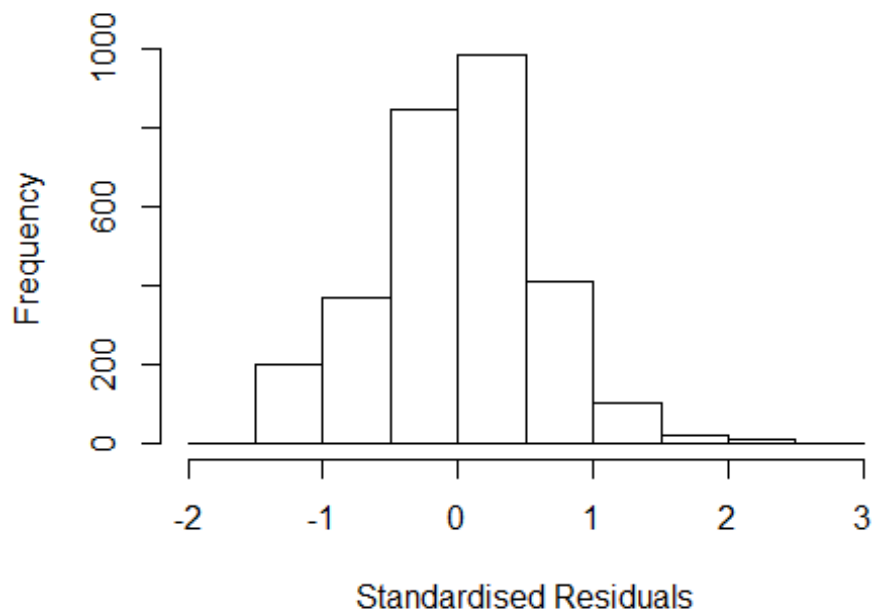


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.27, df = 1, p-value = 0.6
```



```
## [1] "Female first author team size 2018 geometric mean: 1.2774902413674"
## [1] "Male first author team size 2018 geometric mean: 1.12369253387868"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 36000, p-value = 9e-04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.23296644575006"
## [1] "Male last author team size 2018 geometric mean: 1.15934317696147"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34000, p-value = 0.09
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.733 1      1.653
## LastAuthorFemale 2.752 1      1.659
## UniqueAuthors    1.095 4      1.011
## Year              1.139 16     1.004
```

## Residuals from first and last author and team size



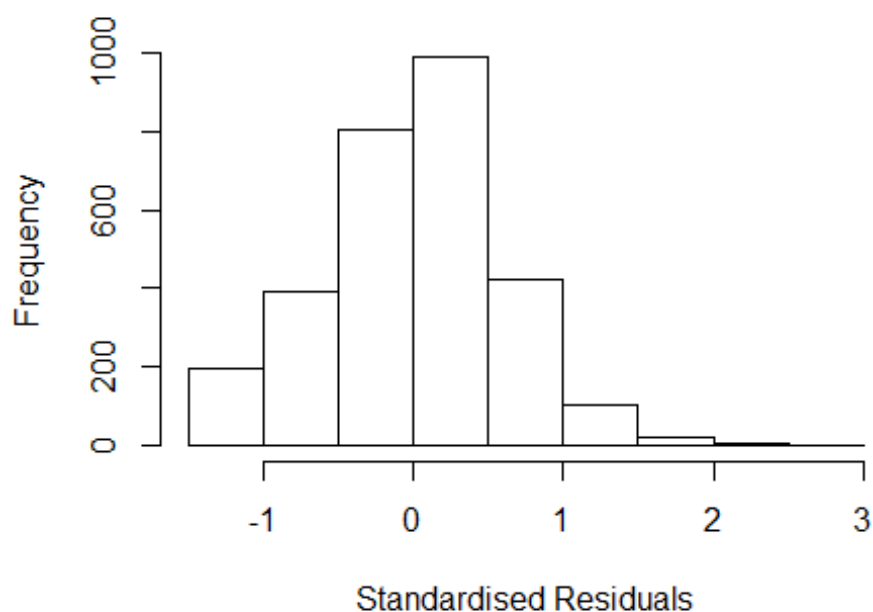
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3984 84857747100 4.135 2012      1201      3      2.664
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5057 -0.3769  0.0273  0.3689  2.6643
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.1273    0.0608   18.53 < 2e-16 ***
## FirstAuthorFemale1  0.0046    0.0365    0.13  0.900
## LastAuthorFemale1  0.0605    0.0366    1.65  0.099 .
## UniqueAuthors2    0.2441    0.0322    7.57 4.9e-14 ***
## UniqueAuthors3    0.2981    0.0451    6.61 4.5e-11 ***
## UniqueAuthors4    0.1712    0.0834    2.05  0.040 *
## UniqueAuthors5    0.1341    0.1262    1.06  0.288
## Year1997         -0.1384    0.0899   -1.54  0.124
## Year1998         -0.1618    0.0840   -1.93  0.054 .
## Year1999         -0.0271    0.0769   -0.35  0.724
```

```

## Year2000          0.0344      0.0758      0.45      0.650
## Year2001         -0.0843      0.0732     -1.15      0.249
## Year2002          0.1323      0.0774      1.71      0.087 .
## Year2003         -0.0322      0.0814     -0.40      0.692
## Year2004          0.0692      0.0759      0.91      0.362
## Year2005          0.1225      0.0774      1.58      0.113
## Year2006          0.0300      0.0713      0.42      0.674
## Year2007         -0.0561      0.0700     -0.80      0.423
## Year2008         -0.0142      0.0714     -0.20      0.842
## Year2009         -0.0416      0.0722     -0.58      0.564
## Year2010         -0.0497      0.0710     -0.70      0.484
## Year2011         -0.0538      0.0719     -0.75      0.455
## Year2012          0.0454      0.0716      0.63      0.526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0421, Adjusted R-squared:  0.0349
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2803 is an outlier with |weight| = 0 ( < 3.4e-05);
## 238 weights are ~= 1. The remaining 2694 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0123 0.8640 0.9520 0.8990 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.41e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.147 1 1.774
## LastAuthorFemale 3.162 1 1.778
## Year 1.057 16 1.002

```

## Residuals from first and last author

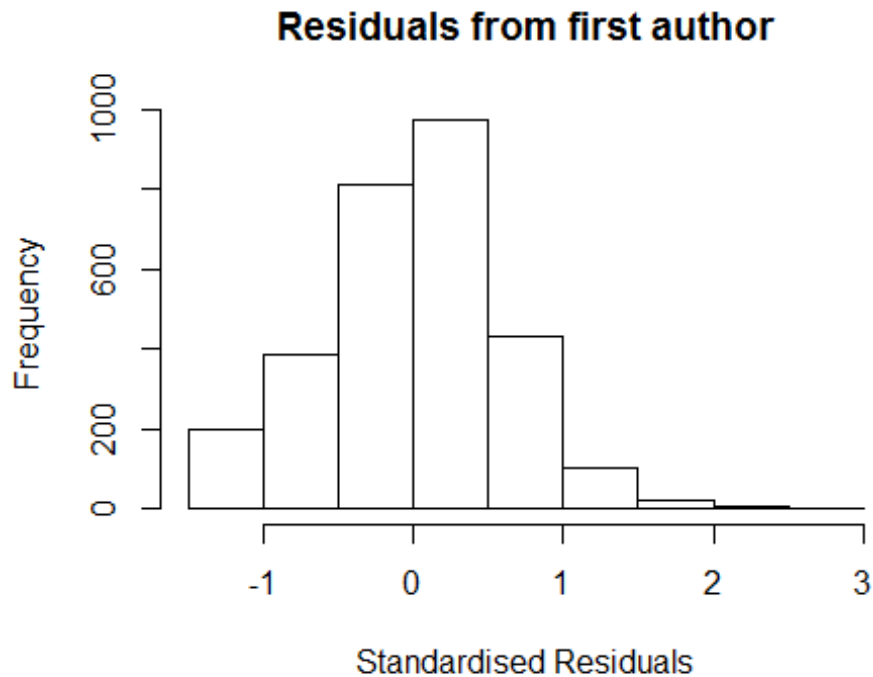


```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3984 84857747100 4.135 2012      1201      3      2.918
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.374 -0.385  0.033  0.379  2.918
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.153012   0.060316   19.12  <2e-16 ***
## FirstAuthorFemale1 0.029077   0.039903    0.73   0.466
## LastAuthorFemale1 0.045194   0.040139    1.13   0.260
## Year1997        -0.120822   0.093168   -1.30   0.195
## Year1998        -0.154392   0.083926   -1.84   0.066 .
## Year1999        -0.014324   0.078306   -0.18   0.855
## Year2000         0.033656   0.075081    0.45   0.654
## Year2001        -0.076528   0.072945   -1.05   0.294
## Year2002         0.146953   0.076891    1.91   0.056 .
## Year2003        -0.034983   0.082255   -0.43   0.671
## Year2004         0.081885   0.075610    1.08   0.279
## Year2005         0.144561   0.077243    1.87   0.061 .
```

```

## Year2006          0.028265    0.071279    0.40    0.692
## Year2007          -0.030692    0.069728   -0.44    0.660
## Year2008          -0.000489    0.071178   -0.01    0.995
## Year2009          -0.030793    0.071982   -0.43    0.669
## Year2010          -0.031032    0.070451   -0.44    0.660
## Year2011          -0.034760    0.071606   -0.49    0.627
## Year2012          0.063737    0.071190    0.90    0.371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0175, Adjusted R-squared:  0.0114
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2803 is an outlier with |weight| = 0 ( < 3.4e-05);
## 237 weights are ~= 1. The remaining 2695 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0302 0.8700 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.41e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.038 1          1.019
## Year              1.038 16          1.001

```



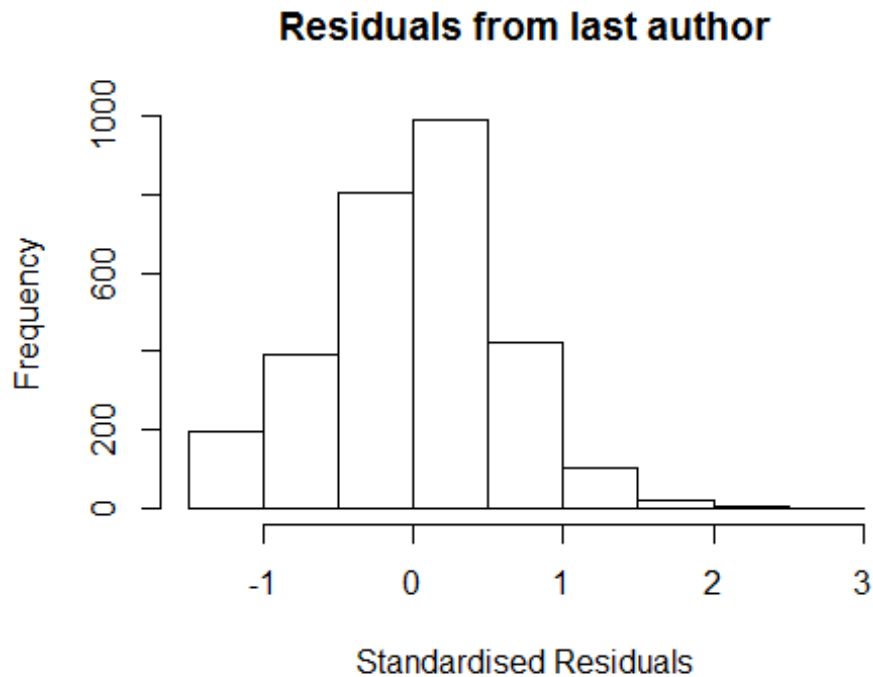
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3984 84857747100 4.135 2012    1201      3      2.918
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3683 -0.3840  0.0326  0.3776  2.9158
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.155319   0.060090   19.23  <2e-16 ***
## FirstAuthorFemale1 0.066664   0.022914    2.91  0.0036 **
## Year1997      -0.119842   0.093504   -1.28  0.2001
## Year1998      -0.155376   0.083864   -1.85  0.0640 .
## Year1999      -0.015302   0.078040   -0.20  0.8446
## Year2000       0.034751   0.075088    0.46  0.6435
## Year2001      -0.075908   0.072822   -1.04  0.2973
## Year2002       0.146292   0.076753    1.91  0.0567 .
## Year2003      -0.035098   0.082092   -0.43  0.6690
## Year2004       0.081006   0.075495    1.07  0.2834
## Year2005       0.146064   0.077076    1.90  0.0582 .
## Year2006       0.028802   0.071142    0.40  0.6856
```



```

## Year2007          -0.029901    0.069638    -0.43    0.6677
## Year2008          -0.000988    0.071097    -0.01    0.9889
## Year2009          -0.030514    0.071887    -0.42    0.6713
## Year2010          -0.029704    0.070308    -0.42    0.6727
## Year2011          -0.033794    0.071436    -0.47    0.6362
## Year2012           0.063891    0.071092     0.90    0.3689
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.017, Adjusted R-squared:  0.0113
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2803 is an outlier with |weight| = 0 ( < 3.4e-05);
## 236 weights are ~= 1. The remaining 2696 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0327 0.8700 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          3.41e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.043 1          1.021
## Year            1.043 16          1.001

```



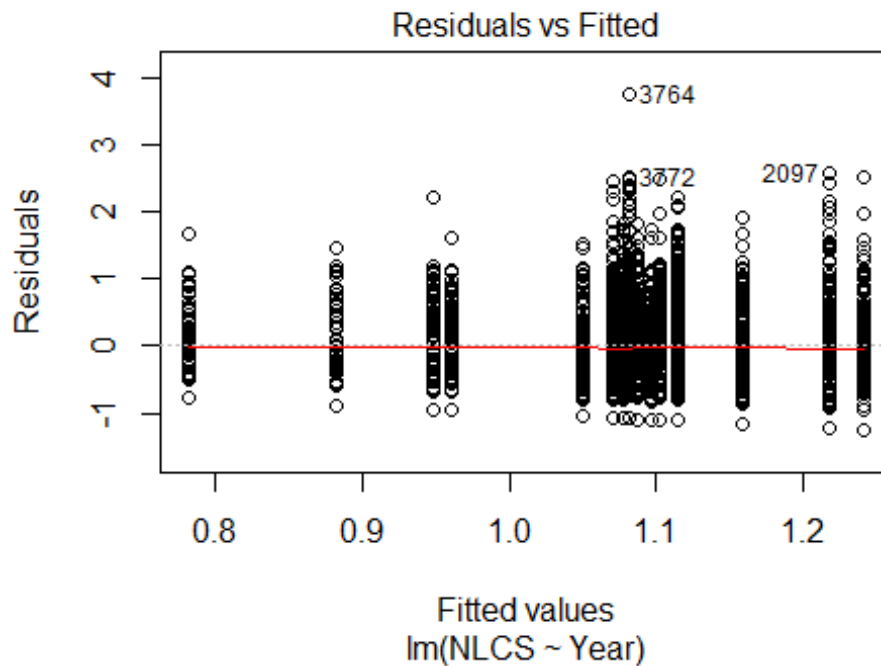
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 3984 84857747100 4.135 2012      1201      3      2.918
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3719 -0.3854  0.0334  0.3781  2.9154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.154757   0.060307   19.15  <2e-16 ***
## LastAuthorFemale1 0.069617   0.023043    3.02  0.0025 **
## Year1997       -0.120472   0.092948   -1.30  0.1950
## Year1998       -0.154135   0.083950   -1.84  0.0665 .
## Year1999       -0.013334   0.078361   -0.17  0.8649
## Year2000        0.034197   0.075068    0.46  0.6487
## Year2001       -0.076734   0.073003   -1.05  0.2933
## Year2002        0.147547   0.076913    1.92  0.0552 .
## Year2003       -0.034826   0.082302   -0.42  0.6722
## Year2004        0.082471   0.075622    1.09  0.2756
## Year2005        0.144950   0.077245    1.88  0.0607 .
## Year2006        0.028677   0.071311    0.40  0.6876
```

```

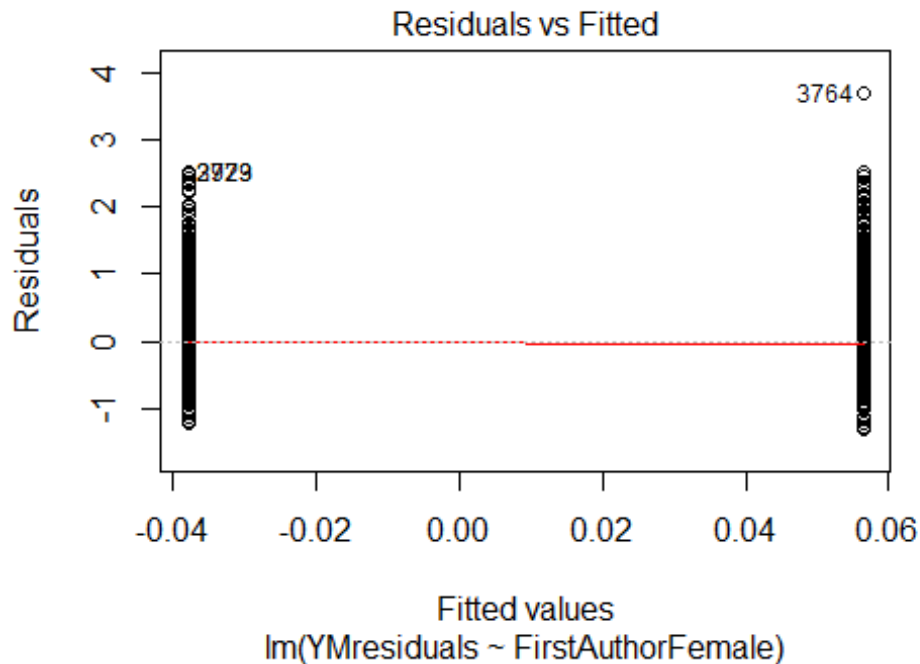
## Year2007      -0.031092    0.069759   -0.45    0.6558
## Year2008      0.000167    0.071172    0.00    0.9981
## Year2009     -0.030384    0.071983   -0.42    0.6730
## Year2010     -0.030883    0.070479   -0.44    0.6613
## Year2011     -0.034548    0.071680   -0.48    0.6299
## Year2012      0.064833    0.071159    0.91    0.3623
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.569
## Multiple R-squared:  0.0173, Adjusted R-squared:  0.0116
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2803 is an outlier with |weight| = 0 ( < 3.4e-05);
## 232 weights are ~= 1. The remaining 2700 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0304 0.8690 0.9500 0.9000 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.41e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2933"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3315"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   80   93   97  101  148  152  144  136  149  210  260  343  286  259  340
## 2011 2012
##  392  516
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   70   81   89   92  118  127  126  116  135  182  227  308  253  224  301

```

```
## 2011 2012
## 337 450
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 70 81 88 92 112 124 122 114 126 178 222 297 246 219 294
## 2011 2012
## 330 433
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 74, df = 16, p-value = 2e-09
```

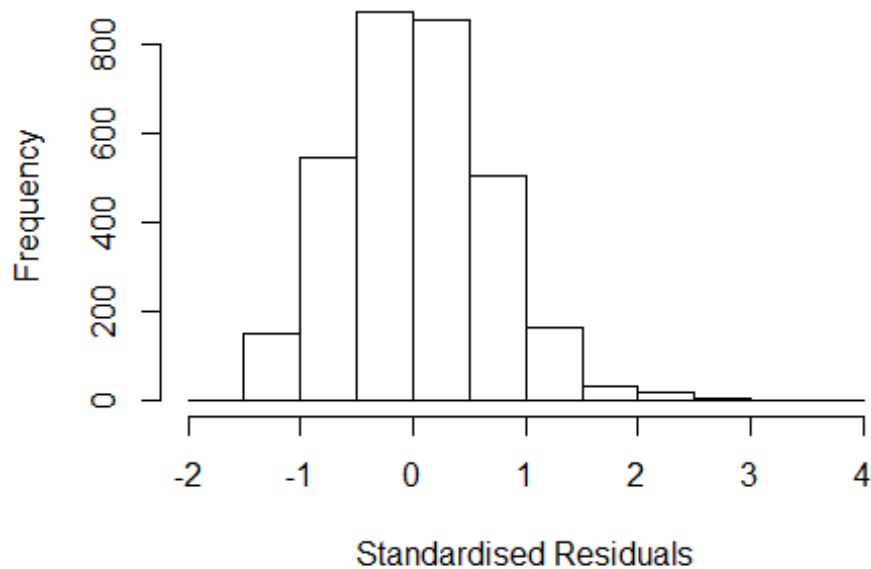


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 9.4, df = 1, p-value = 0.002
```



```
## [1] "Female first author team size 2018 geometric mean: 1.33131489037301"
## [1] "Male first author team size 2018 geometric mean: 1.31978961284962"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 32000, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.36633246459063"
## [1] "Male last author team size 2018 geometric mean: 1.2849350002711"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 34000, p-value = 0.07
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.377 1      1.542
## LastAuthorFemale  2.347 1      1.532
## UniqueAuthors    1.074 4      1.009
## Year              1.107 16     1.003
```

## Residuals from first and last author and team size



```
## [1] "List of 5 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1051  5044246868 3.743 2004    1213     3     2.554
## 2097  57749139264 3.783 2008    1213     3     2.637
## 3764  84861974217 4.826 2012    3309     2     3.552
## 3767  84862121724 3.419 2012    1203     2     2.519
## 3772  84861566134 3.608 2012    1213     3     2.619
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.54126 -0.44091  0.00129  0.46119  3.55218
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7909    0.0746   10.60 < 2e-16 ***
## FirstAuthorFemale1 0.1019    0.0382    2.67  0.0077 **
## LastAuthorFemale1 -0.0128    0.0382   -0.33  0.7383
## UniqueAuthors2    0.2850    0.0347    8.21 3.2e-16 ***
## UniqueAuthors3    0.2605    0.0437    5.96 2.8e-09 ***
## UniqueAuthors4    0.2308    0.0879    2.62  0.0087 **
```

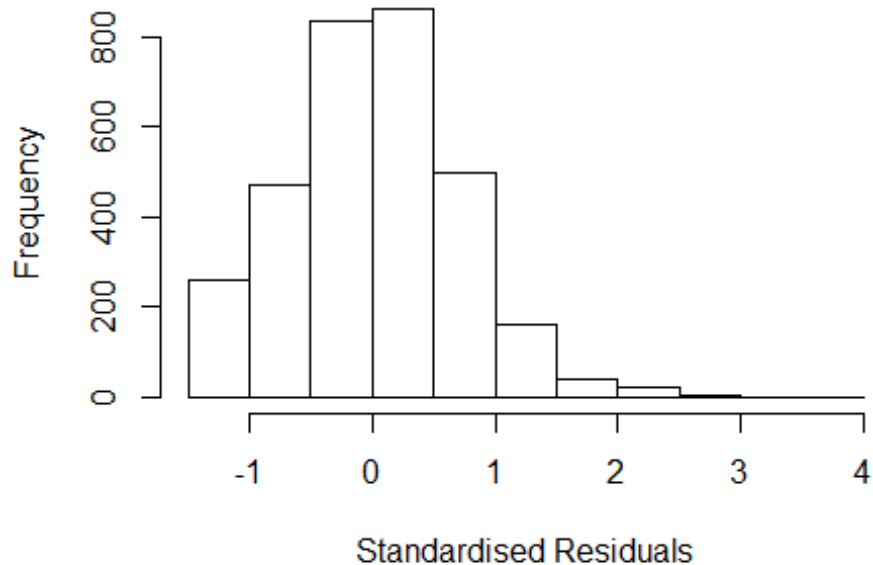
```

## UniqueAuthors5      0.5524      0.1164      4.75 2.2e-06 ***
## Year1997             -0.1010      0.0972     -1.04 0.2989
## Year1998             0.0302      0.0998      0.30 0.7624
## Year1999             0.0696      0.0980      0.71 0.4776
## Year2000             0.1329      0.0944      1.41 0.1592
## Year2001             0.1913      0.0912      2.10 0.0360 *
## Year2002             0.2038      0.0992      2.05 0.0400 *
## Year2003             0.1022      0.0978      1.04 0.2962
## Year2004             0.3090      0.0954      3.24 0.0012 **
## Year2005             0.2640      0.0884      2.99 0.0028 **
## Year2006             0.2024      0.0867      2.33 0.0196 *
## Year2007             0.1431      0.0827      1.73 0.0837 .
## Year2008             0.2664      0.0854      3.12 0.0018 **
## Year2009             0.1615      0.0873      1.85 0.0642 .
## Year2010             0.1809      0.0827      2.19 0.0288 *
## Year2011             0.1886      0.0851      2.22 0.0267 *
## Year2012             0.1088      0.0841      1.29 0.1957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.668
## Multiple R-squared:  0.0559, Adjusted R-squared:  0.0493
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2933 is an outlier with |weight| = 0 ( < 3.2e-05);
## 266 weights are ~= 1. The remaining 2881 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0845 0.8620 0.9510 0.9090 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.18e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.407 1          1.551

```

```
## LastAuthorFemale 2.391 1 1.546
## Year 1.031 16 1.001
```

### Residuals from first and last author



```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2097 57749139264 3.783 2008    1213     3    2.561
## 3729 84864030671 3.567 2012    1208     2    2.618
## 3764 84861974217 4.826 2012    3309     2    3.776
## 3772 84861566134 3.608 2012    1213     3    2.558
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26068 -0.46223  0.00745  0.46423  3.77613
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8428    0.0774   10.88 < 2e-16 ***
## FirstAuthorFemale1 0.1013    0.0393    2.58 0.01000 **
## LastAuthorFemale1 -0.0173    0.0393   -0.44 0.65956
## Year1997        -0.0970    0.0996   -0.97 0.33038
## Year1998         0.0481    0.1031    0.47 0.64085
## Year1999         0.0834    0.1030    0.81 0.41810
```

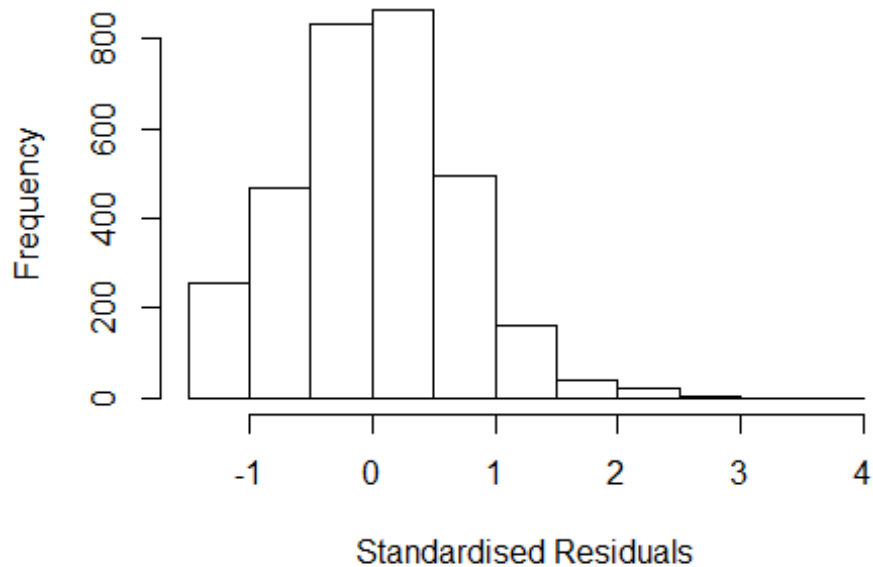


```

## Year2000      0.1480      0.0959      1.54  0.12266
## Year2001      0.2073      0.0937      2.21  0.02703 *
## Year2002      0.2005      0.1025      1.96  0.05045 .
## Year2003      0.1572      0.1017      1.54  0.12256
## Year2004      0.3340      0.0985      3.39  0.00071 ***
## Year2005      0.2887      0.0918      3.15  0.00168 **
## Year2006      0.2234      0.0898      2.49  0.01285 *
## Year2007      0.1860      0.0859      2.17  0.03033 *
## Year2008      0.2956      0.0885      3.34  0.00084 ***
## Year2009      0.2001      0.0906      2.21  0.02731 *
## Year2010      0.2137      0.0857      2.49  0.01267 *
## Year2011      0.2185      0.0880      2.48  0.01308 *
## Year2012      0.1231      0.0874      1.41  0.15872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.02,   Adjusted R-squared:  0.0144
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2933 is an outlier with |weight| = 0 ( < 3.2e-05);
## 253 weights are ~= 1. The remaining 2894 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.110  0.870  0.952  0.910  0.985  0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           3.18e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.024 1           1.012
## Year           1.024 16           1.001

```

## Residuals from first author



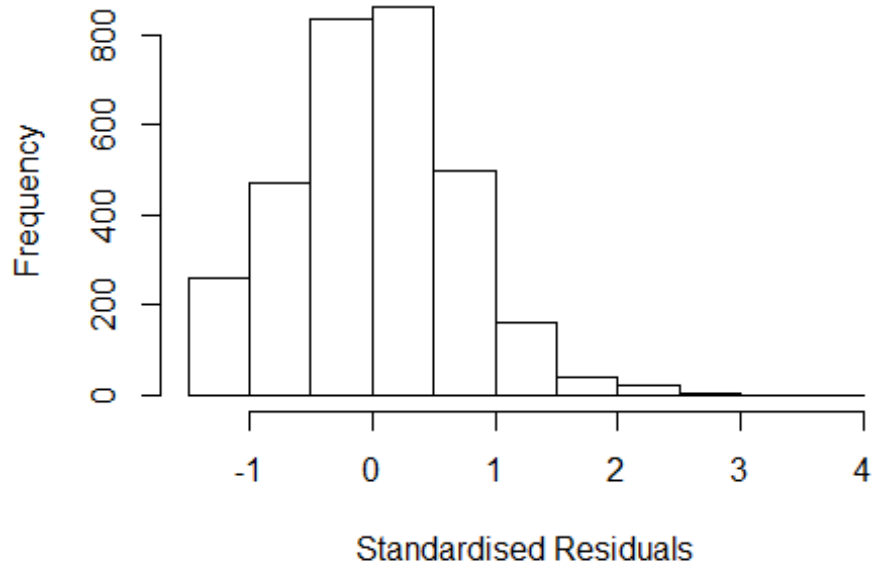
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2097 57749139264 3.783 2008    1213      3    2.561
## 3729 84864030671 3.567 2012    1208      2    2.618
## 3764 84861974217 4.826 2012    3309      2    3.776
## 3772 84861566134 3.608 2012    1213      3    2.558
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26220 -0.46044  0.00673  0.45996  3.77335
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8417    0.0774   10.88 < 2e-16 ***
## FirstAuthorFemale1 0.0880    0.0256    3.43 0.00061 ***
## Year1997      -0.0969    0.0995   -0.97 0.33044
## Year1998       0.0474    0.1031    0.46 0.64587
## Year1999       0.0836    0.1030    0.81 0.41729
## Year2000       0.1478    0.0957    1.54 0.12275
## Year2001       0.2064    0.0937    2.20 0.02764 *
## Year2002       0.1998    0.1024    1.95 0.05116 .
## Year2003       0.1572    0.1017    1.55 0.12230
```

```

## Year2004          0.3325      0.0984      3.38  0.00074 ***
## Year2005          0.2882      0.0917      3.14  0.00170 **
## Year2006          0.2227      0.0897      2.48  0.01305 *
## Year2007          0.1856      0.0858      2.16  0.03062 *
## Year2008          0.2948      0.0884      3.33  0.00086 ***
## Year2009          0.2000      0.0906      2.21  0.02735 *
## Year2010          0.2126      0.0856      2.48  0.01304 *
## Year2011          0.2186      0.0880      2.48  0.01303 *
## Year2012          0.1230      0.0873      1.41  0.15910
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.0199, Adjusted R-squared:  0.0146
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 2933 is an outlier with |weight| = 0 ( < 3.2e-05);
## 254 weights are ~= 1. The remaining 2893 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.116  0.870   0.952   0.910   0.985   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.18e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.017 1           1.009
## Year             1.017 16           1.001

```

## Residuals from last author



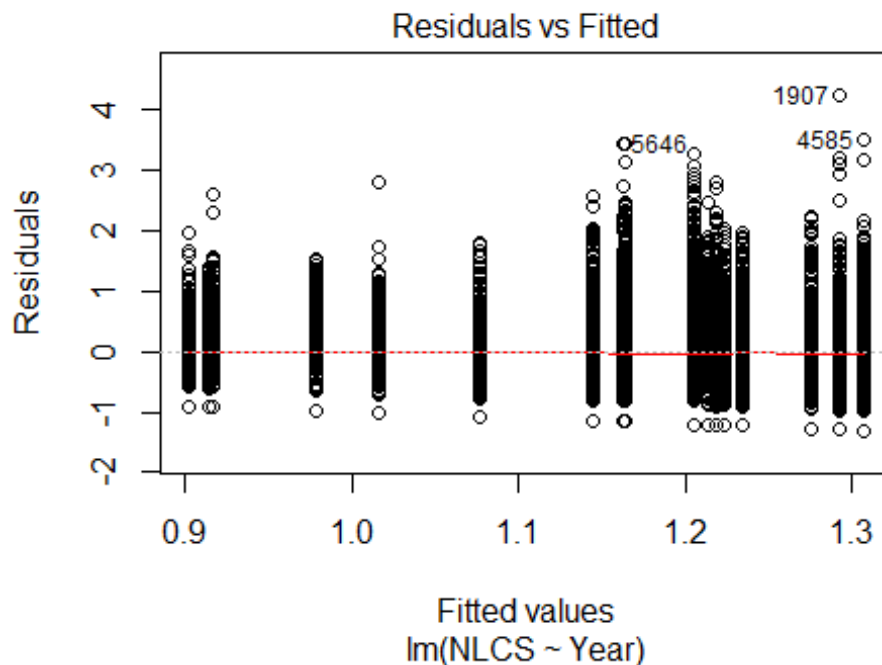
```
## [1] "List of 4 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 2097 57749139264 3.783 2008    1213     3    2.561
## 3729 84864030671 3.567 2012    1208     2    2.618
## 3764 84861974217 4.826 2012    3309     2    3.776
## 3772 84861566134 3.608 2012    1213     3    2.558
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.24324 -0.46825  0.00491  0.46114  3.78696
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8495    0.0775   10.96 < 2e-16 ***
## LastAuthorFemale1 0.0608    0.0256    2.37  0.01775 *
## Year1997       -0.0988    0.0994   -0.99  0.32035
## Year1998        0.0487    0.1036    0.47  0.63796
## Year1999        0.0853    0.1030    0.83  0.40781
## Year2000        0.1506    0.0959    1.57  0.11622
## Year2001        0.2103    0.0942    2.23  0.02563 *
## Year2002        0.2029    0.1024    1.98  0.04769 *
## Year2003        0.1632    0.1019    1.60  0.10920
```

```

## Year2004          0.3329      0.0983      3.39  0.00072 ***
## Year2005          0.2906      0.0917      3.17  0.00155 **
## Year2006          0.2228      0.0897      2.48  0.01311 *
## Year2007          0.1886      0.0860      2.19  0.02839 *
## Year2008          0.2972      0.0885      3.36  0.00079 ***
## Year2009          0.2057      0.0908      2.27  0.02350 *
## Year2010          0.2151      0.0858      2.51  0.01225 *
## Year2011          0.2231      0.0882      2.53  0.01153 *
## Year2012          0.1287      0.0874      1.47  0.14084
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.684
## Multiple R-squared:  0.0179, Adjusted R-squared:  0.0125
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## observation 2933 is an outlier with |weight| = 0 ( < 3.2e-05);
## 254 weights are ~ = 1. The remaining 2893 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.125  0.869  0.951   0.910  0.985   0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          3.18e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3148"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3316"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 246 238 254 285 290 302 397 368 410 444 552 574 560 604 859
## 2011 2012
## 1102 1197

```

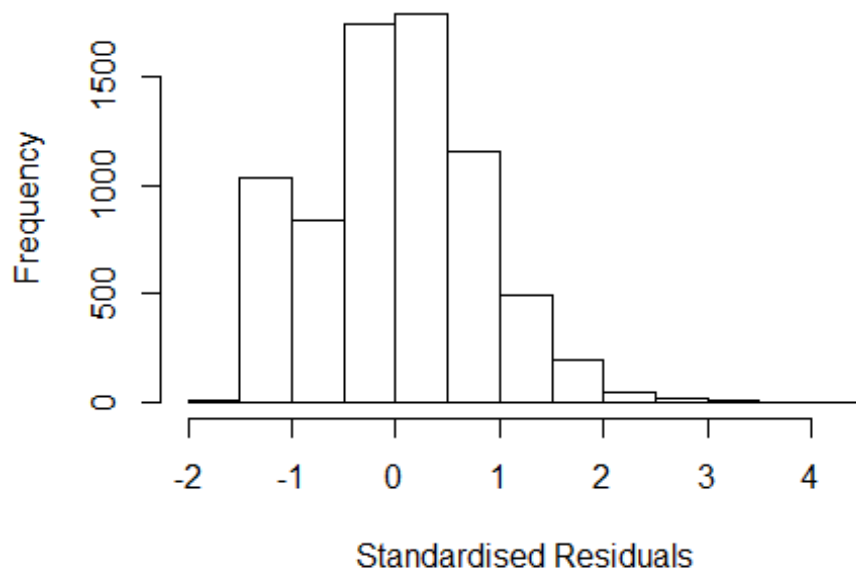
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 223 194 217 253 243 235 348 330 353 376 478 492 478 522 748
## 2011 2012
## 953 1027
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 222 190 216 249 241 232 344 322 350 370 473 485 472 517 730
## 2011 2012
## 925 1006
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 130, df = 16, p-value <2e-16
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.066, df = 1, p-value = 0.8
```



## Residuals from first and last author and team size



```
## [1] "List of 29 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 762  1542307199 3.506 1999    3316      1    2.630
## 1106 61249468539 3.822 2000    1208      3    2.858
## 1825 1642297599 4.387 2002    1213      2    2.777
## 1874 48949111059 4.224 2002    1208      2    3.020
## 1905 84997942206 4.487 2002    1213      2    3.284
## 1906 84998080695 3.794 2002    1213      2    2.591
## 1907 84998183113 5.529 2002    1213      2    4.326
## 2742 84998183199 3.693 2004    1213      2    2.521
## 4206 34250156387 4.006 2007    3316      1    2.865
## 4232 34250648363 3.919 2007    1213      2    2.777
## 4585 60950457792 4.816 2008    1213      2    3.591
## 5006 70449388215 4.476 2008    1213      2    2.868
## 5264 77956345467 3.902 2009    3316      1    2.814
## 5646 70449753866 4.612 2009    1213      2    3.525
## 6403 77949526431 3.723 2010    3316      1    2.662
## 6748 84855915732 4.594 2011    1208      4    3.529
## 6749 84855916060 4.298 2011    1208      4    3.232
## 6750 84855927603 3.619 2011    1208      4    2.554
## 7194 79959538449 3.615 2011    3316      1    2.549
## 7613 79952723945 3.604 2011    1213      3    2.538
## 8346 84865966222 3.775 2012    3316      1    2.677
## 8452 84860189956 3.633 2012    1201      3    2.536
## 8545 84861566134 3.608 2012    1213      3    2.510
## 8687 84857747100 4.135 2012    1201      3    2.611
## 8707 84859894535 4.283 2012    1202      3    2.805
```



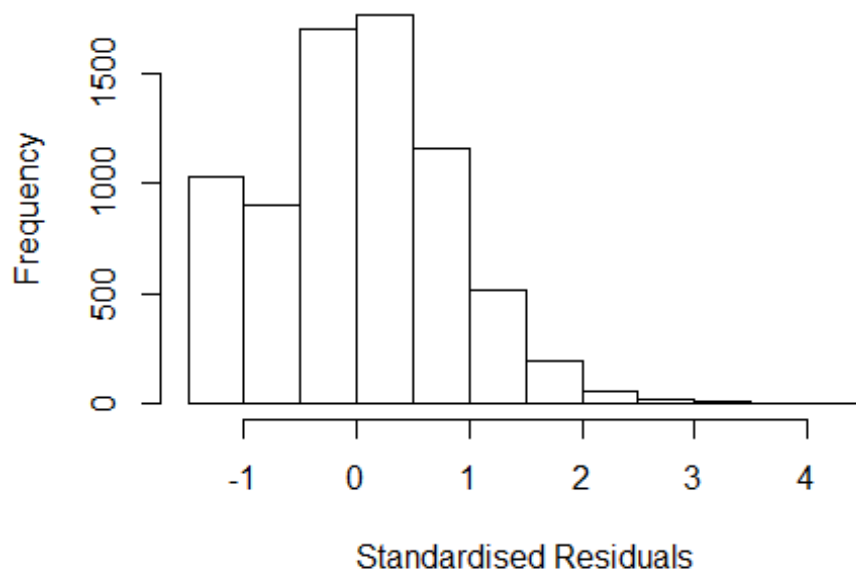
```

## 8872 84857213873 3.996 2012      1202      2      2.899
## 9009 84856425812 3.956 2012      1211      3      2.859
## 9010 84856433228 4.077 2012      1211      3      2.624
## 9011 84856451242 4.492 2012      1211      3      3.394
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6429 -0.5233  0.0102  0.5261  4.3262
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9266      0.0484   19.14 < 2e-16 ***
## FirstAuthorFemale1 0.0256      0.0379    0.68 0.49876
## LastAuthorFemale1 -0.0243      0.0379   -0.64 0.52116
## UniqueAuthors2     0.3813      0.0324   11.76 < 2e-16 ***
## UniqueAuthors3     0.4279      0.0589    7.27 4.1e-13 ***
## UniqueAuthors4     0.4834      0.1098    4.40 1.1e-05 ***
## UniqueAuthors5     0.4664      0.1231    3.79 0.00015 ***
## Year1997          -0.0707      0.0702   -1.01 0.31396
## Year1998          -0.0460      0.0684   -0.67 0.50100
## Year1999          -0.0515      0.0656   -0.78 0.43262
## Year2000           0.0374      0.0662    0.57 0.57165
## Year2001           0.0867      0.0651    1.33 0.18315
## Year2002           0.2762      0.0661    4.18 2.9e-05 ***
## Year2003           0.2266      0.0643    3.53 0.00043 ***
## Year2004           0.2457      0.0642    3.83 0.00013 ***
## Year2005           0.2884      0.0656    4.39 1.1e-05 ***
## Year2006           0.2487      0.0594    4.18 2.9e-05 ***
## Year2007           0.2144      0.0601    3.57 0.00036 ***
## Year2008           0.2987      0.0608    4.91 9.3e-07 ***
## Year2009           0.1604      0.0583    2.75 0.00600 **
## Year2010           0.1334      0.0558    2.39 0.01693 *
## Year2011           0.1379      0.0557    2.48 0.01329 *
## Year2012           0.1699      0.0565    3.01 0.00264 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.8
## Multiple R-squared:  0.041, Adjusted R-squared:  0.0382
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1551 is an outlier with |weight| = 0 ( < 1.4e-05);

```

```
## 628 weights are ~= 1. The remaining 6715 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0066 0.8460 0.9520 0.9120 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.572 1 2.138
## LastAuthorFemale 4.582 1 2.141
## Year 1.022 16 1.001
```

## Residuals from first and last author



```

## [1] "List of 26 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 762   1542307199 3.506 1999    3316      1    2.602
## 1106  61249468539 3.822 2000    1208      3    2.827
## 1825  1642297599 4.387 2002    1213      2    3.109
## 1874  48949111059 4.224 2002    1208      2    2.973
## 1905  84997942206 4.487 2002    1213      2    3.244
## 1906  84998080695 3.794 2002    1213      2    2.551
## 1907  84998183113 5.529 2002    1213      2    4.286
## 4206  34250156387 4.006 2007    3316      1    2.820
## 4232  34250648363 3.919 2007    1213      2    2.725
## 4585  60950457792 4.816 2008    1213      2    3.545
## 5006  70449388215 4.476 2008    1213      2    3.197
## 5264  77956345467 3.902 2009    3316      1    2.756
## 5646  70449753866 4.612 2009    1213      2    3.474
## 6403  77949526431 3.723 2010    3316      1    2.603
## 6748  84855915732 4.594 2011    1208      4    3.480
## 6749  84855916060 4.298 2011    1208      4    3.176
## 6750  84855927603 3.619 2011    1208      4    2.505
## 8343  84863612069 3.775 2012    3316      1    2.631
## 8345  84863632617 3.839 2012    3316      1    2.695
## 8346  84865966222 3.775 2012    3316      1    2.623
## 8687  84857747100 4.135 2012    1201      3    2.991
## 8707  84859894535 4.283 2012    1202      3    3.139
## 8872  84857213873 3.996 2012    1202      2    2.852
## 9009  84856425812 3.956 2012    1211      3    2.812
## 9010  84856433228 4.077 2012    1211      3    2.960
## 9011  84856451242 4.492 2012    1211      3    3.340
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2786 -0.5270  0.0122  0.5370  4.2860
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9654    0.0483   20.00 < 2e-16 ***
## FirstAuthorFemale1 0.0351    0.0417    0.84  0.39998
## LastAuthorFemale1 -0.0272    0.0418   -0.65  0.51489
## Year1997        -0.0818    0.0702   -1.17  0.24401
## Year1998        -0.0599    0.0682   -0.88  0.37970
## Year1999        -0.0690    0.0657   -1.05  0.29353
## Year2000         0.0300    0.0663    0.45  0.65078
## Year2001         0.0902    0.0654    1.38  0.16806
## Year2002         0.2776    0.0658    4.22  2.5e-05 ***
## Year2003         0.2310    0.0653    3.54  0.00041 ***

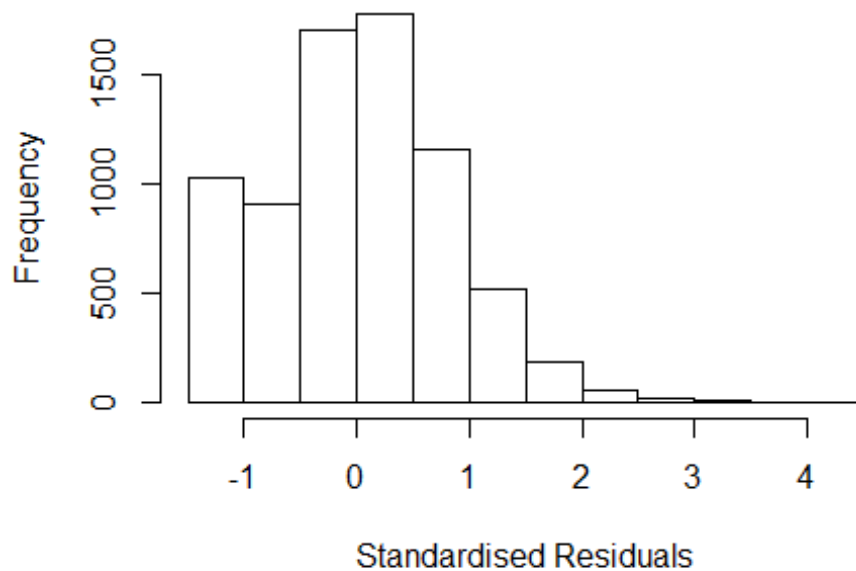
```

```

## Year2004          0.2336      0.0643      3.63  0.00028 ***
## Year2005          0.2887      0.0658      4.39  1.1e-05 ***
## Year2006          0.2537      0.0598      4.24  2.2e-05 ***
## Year2007          0.2207      0.0604      3.65  0.00026 ***
## Year2008          0.3053      0.0613      4.98  6.5e-07 ***
## Year2009          0.1723      0.0586      2.94  0.00329 **
## Year2010          0.1472      0.0558      2.64  0.00838 **
## Year2011          0.1486      0.0559      2.66  0.00786 **
## Year2012          0.1783      0.0567      3.15  0.00166 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.81
## Multiple R-squared:  0.0165, Adjusted R-squared:  0.014
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1551 is an outlier with |weight| = 0 ( < 1.4e-05);
## 616 weights are ~= 1. The remaining 6727 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0163  0.8490  0.9520  0.9120  0.9860  0.9990
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          1.36e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.016  1           1.008
## Year              1.016 16           1.000

```

## Residuals from first author



```
## [1] "List of 26 outliers with residuals above 2.5"
##           ScopusId  NLCS Year OneField Fields residuals
## 762    1542307199 3.506 1999    3316      1    2.602
## 1106   61249468539 3.822 2000    1208      3    2.827
## 1825   1642297599 4.387 2002    1213      2    3.109
## 1874   48949111059 4.224 2002    1208      2    2.973
## 1905   84997942206 4.487 2002    1213      2    3.244
## 1906   84998080695 3.794 2002    1213      2    2.551
## 1907   84998183113 5.529 2002    1213      2    4.286
## 4206   34250156387 4.006 2007    3316      1    2.820
## 4232   34250648363 3.919 2007    1213      2    2.725
## 4585   60950457792 4.816 2008    1213      2    3.545
## 5006   70449388215 4.476 2008    1213      2    3.197
## 5264   77956345467 3.902 2009    3316      1    2.756
## 5646   70449753866 4.612 2009    1213      2    3.474
## 6403   77949526431 3.723 2010    3316      1    2.603
## 6748   84855915732 4.594 2011    1208      4    3.480
## 6749   84855916060 4.298 2011    1208      4    3.176
## 6750   84855927603 3.619 2011    1208      4    2.505
## 8343   84863612069 3.775 2012    3316      1    2.631
## 8345   84863632617 3.839 2012    3316      1    2.695
## 8346   84865966222 3.775 2012    3316      1    2.623
## 8687   84857747100 4.135 2012    1201      3    2.991
## 8707   84859894535 4.283 2012    1202      3    3.139
## 8872   84857213873 3.996 2012    1202      2    2.852
## 9009   84856425812 3.956 2012    1211      3    2.812
## 9010   84856433228 4.077 2012    1211      3    2.960
```

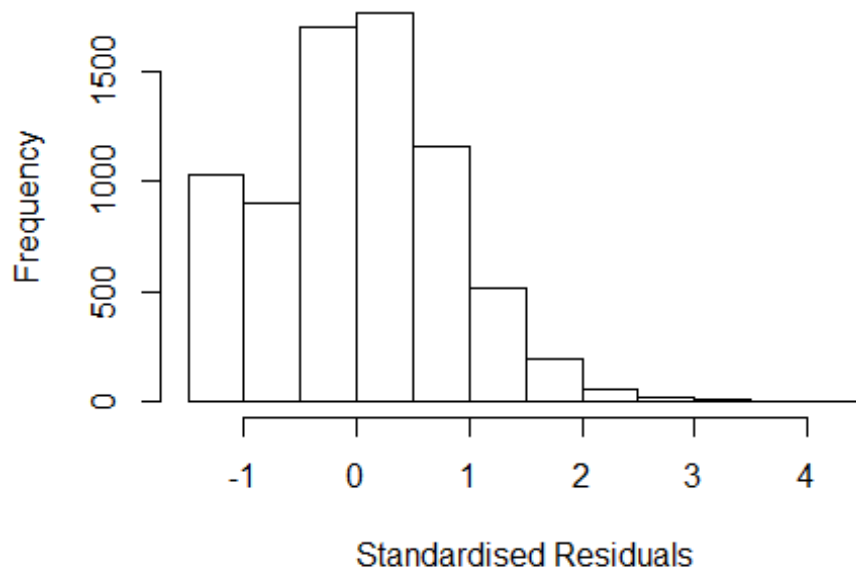
```

## 9011 84856451242 4.492 2012      1211      3      3.340
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2798 -0.5277  0.0133  0.5377  4.2870
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9643     0.0482   20.01 < 2e-16 ***
## FirstAuthorFemale1 0.0107     0.0197    0.54 0.58582
## Year1997        -0.0809     0.0702   -1.15 0.24863
## Year1998        -0.0594     0.0682   -0.87 0.38355
## Year1999        -0.0688     0.0657   -1.05 0.29504
## Year2000         0.0302     0.0663    0.46 0.64906
## Year2001         0.0905     0.0654    1.38 0.16671
## Year2002         0.2777     0.0658    4.22 2.5e-05 ***
## Year2003         0.2311     0.0654    3.54 0.00041 ***
## Year2004         0.2335     0.0643    3.63 0.00028 ***
## Year2005         0.2888     0.0658    4.39 1.2e-05 ***
## Year2006         0.2534     0.0598    4.24 2.3e-05 ***
## Year2007         0.2204     0.0605    3.65 0.00027 ***
## Year2008         0.3047     0.0613    4.97 6.9e-07 ***
## Year2009         0.1727     0.0586    2.95 0.00321 **
## Year2010         0.1471     0.0558    2.64 0.00843 **
## Year2011         0.1486     0.0559    2.66 0.00790 **
## Year2012         0.1779     0.0567    3.14 0.00171 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.81
## Multiple R-squared:  0.0164, Adjusted R-squared:  0.0141
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1551 is an outlier with |weight| = 0 ( < 1.4e-05);
## 619 weights are ~ 1. The remaining 6724 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.016  0.849  0.952  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.36e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01

```

```
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##          500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##          0          1000          0
##          psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
##      compute.outlier.stats
##          "SM"
##      seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.018 1          1.009
## Year              1.018 16          1.001
```

### Residuals from last author



```
## [1] "List of 26 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 762  1542307199 3.506 1999    3316     1    2.602
## 1106 61249468539 3.822 2000    1208     3    2.827
## 1825 1642297599 4.387 2002    1213     2    3.109
## 1874 48949111059 4.224 2002    1208     2    2.973
## 1905 84997942206 4.487 2002    1213     2    3.244
## 1906 84998080695 3.794 2002    1213     2    2.551
## 1907 84998183113 5.529 2002    1213     2    4.286
## 4206 34250156387 4.006 2007    3316     1    2.820
## 4232 34250648363 3.919 2007    1213     2    2.725
## 4585 60950457792 4.816 2008    1213     2    3.545
## 5006 70449388215 4.476 2008    1213     2    3.197
```

```

## 5264 77956345467 3.902 2009      3316      1      2.756
## 5646 70449753866 4.612 2009      1213      2      3.474
## 6403 77949526431 3.723 2010      3316      1      2.603
## 6748 84855915732 4.594 2011      1208      4      3.480
## 6749 84855916060 4.298 2011      1208      4      3.176
## 6750 84855927603 3.619 2011      1208      4      2.505
## 8343 84863612069 3.775 2012      3316      1      2.631
## 8345 84863632617 3.839 2012      3316      1      2.695
## 8346 84865966222 3.775 2012      3316      1      2.623
## 8687 84857747100 4.135 2012      1201      3      2.991
## 8707 84859894535 4.283 2012      1202      3      3.139
## 8872 84857213873 3.996 2012      1202      2      2.852
## 9009 84856425812 3.956 2012      1211      3      2.812
## 9010 84856433228 4.077 2012      1211      3      2.960
## 9011 84856451242 4.492 2012      1211      3      3.340
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2761 -0.5280  0.0122  0.5370  4.2845
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96629    0.04829   20.01 < 2e-16 ***
## LastAuthorFemale1 0.00431    0.01969    0.22  0.82685
## Year1997      -0.08055    0.07016   -1.15  0.25098
## Year1998      -0.05951    0.06816   -0.87  0.38266
## Year1999      -0.06833    0.06563   -1.04  0.29786
## Year2000       0.03022    0.06629    0.46  0.64853
## Year2001       0.09067    0.06541    1.39  0.16575
## Year2002       0.27821    0.06582    4.23  2.4e-05 ***
## Year2003       0.23180    0.06533    3.55  0.00039 ***
## Year2004       0.23365    0.06428    3.63  0.00028 ***
## Year2005       0.28903    0.06577    4.39  1.1e-05 ***
## Year2006       0.25382    0.05979    4.25  2.2e-05 ***
## Year2007       0.22072    0.06044    3.65  0.00026 ***
## Year2008       0.30546    0.06133    4.98  6.5e-07 ***
## Year2009       0.17329    0.05858    2.96  0.00310 **
## Year2010       0.14781    0.05582    2.65  0.00811 **
## Year2011       0.14952    0.05590    2.67  0.00749 **
## Year2012       0.17875    0.05668    3.15  0.00162 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.81
## Multiple R-squared:  0.0164, Adjusted R-squared:  0.0141

```

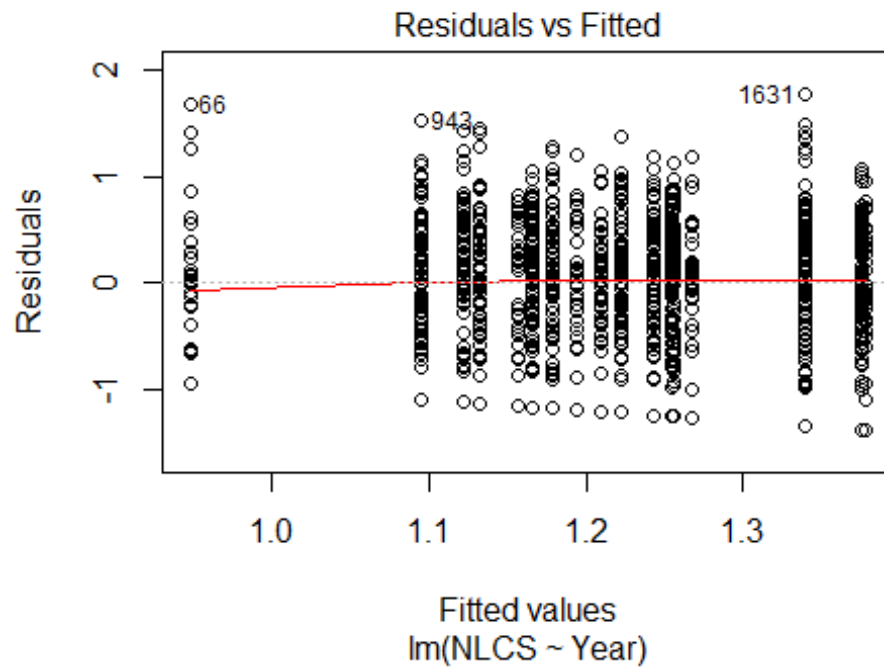


```

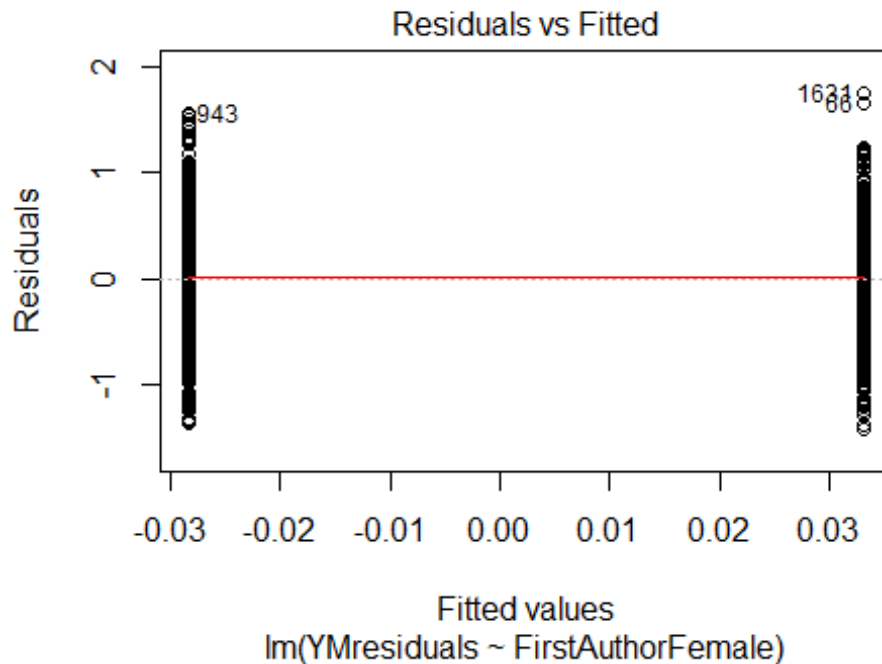
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1551 is an outlier with |weight| = 0 ( < 1.4e-05);
## 627 weights are ~= 1. The remaining 6716 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0163 0.8490 0.9520 0.9120 0.9860 0.9990
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           1.36e-05           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 7344"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3317"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   50  46  71  66  70  82  97  88  80  88  107  94  93  123  109
## 2011 2012
##  132  141
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34  32  44  39  36  31  73  81  63  81  97  81  79  104  94
## 2011 2012
##  116  121
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   32  31  41  39  32  29  72  75  60  81  94  77  73  98  90
## 2011 2012
##  111  113
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances

```

```
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```

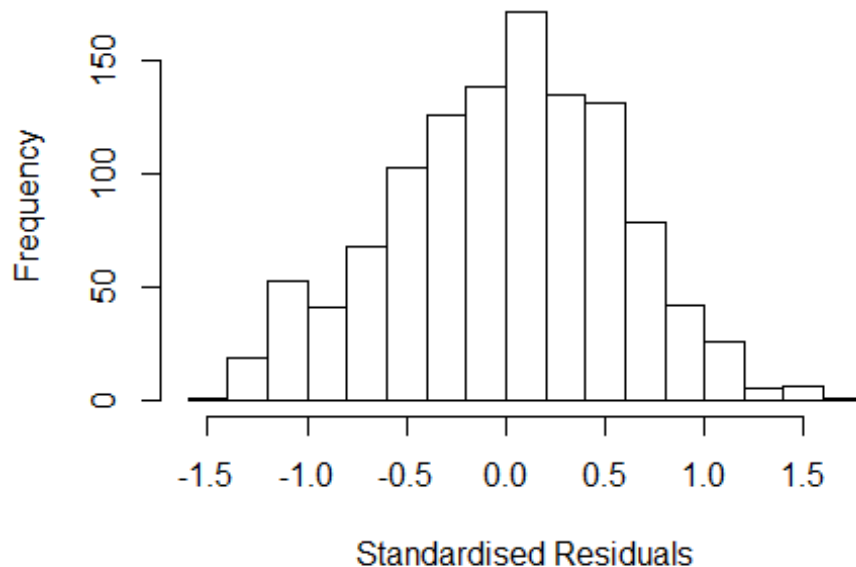


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 5, df = 1, p-value = 0.03
```



```
## [1] "Female first author team size 2018 geometric mean: 1.29291052181876"
## [1] "Male first author team size 2018 geometric mean: 1.26504828681519"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.24007837136781"
## [1] "Male last author team size 2018 geometric mean: 1.34139668039395"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1300, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.338 1      1.529
## LastAuthorFemale 2.329 1      1.526
## UniqueAuthors    1.348 4      1.038
## Year             1.475 16      1.012
```

## Residuals from first and last author and team size



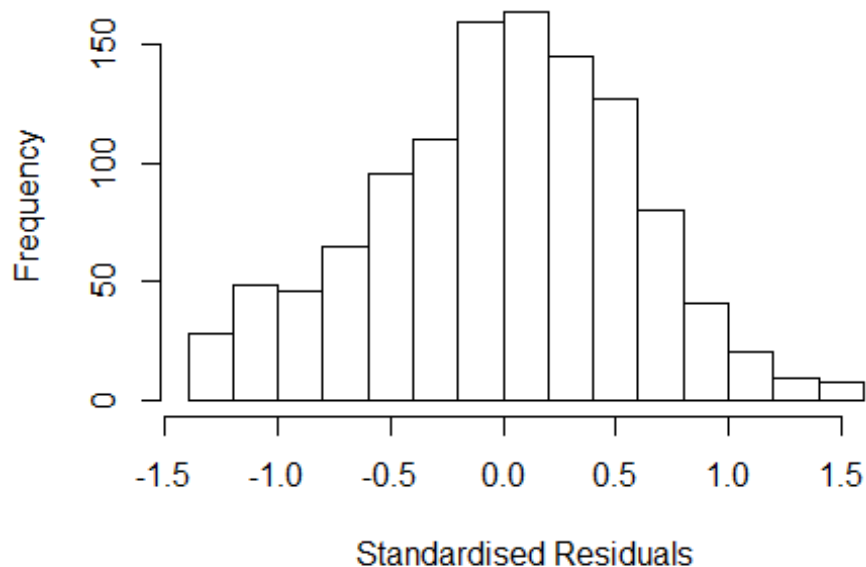
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4351 -0.3923  0.0275  0.4028  1.6402
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.15780    0.12651   9.15 < 2e-16 ***
## FirstAuthorFemale1 -0.00137    0.05270  -0.03  0.97933
## LastAuthorFemale1  0.08331    0.05285   1.58  0.11520
## UniqueAuthors2     0.16049    0.04531   3.54  0.00041 ***
## UniqueAuthors3     0.16123    0.06011   2.68  0.00742 **
## UniqueAuthors4     0.48261    0.09550   5.05  5.1e-07 ***
## UniqueAuthors5     0.42383    0.17374   2.44  0.01487 *
## Year1997          -0.34795    0.16081  -2.16  0.03069 *
## Year1998          -0.06462    0.15568  -0.42  0.67816
## Year1999           0.03045    0.14854   0.20  0.83763
```

```

## Year2000      0.15206      0.16288      0.93  0.35073
## Year2001     -0.06065      0.17255     -0.35  0.72529
## Year2002     -0.11335      0.14455     -0.78  0.43312
## Year2003     -0.11943      0.14674     -0.81  0.41589
## Year2004      0.01615      0.14925      0.11  0.91384
## Year2005      0.04155      0.14060      0.30  0.76764
## Year2006     -0.17496      0.14160     -1.24  0.21685
## Year2007     -0.10695      0.14779     -0.72  0.46943
## Year2008      0.00476      0.14502      0.03  0.97381
## Year2009     -0.01220      0.13729     -0.09  0.92920
## Year2010     -0.06380      0.14016     -0.46  0.64906
## Year2011      0.11743      0.13720      0.86  0.39225
## Year2012      0.04914      0.14217      0.35  0.72968
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.573
## Multiple R-squared:  0.0657, Adjusted R-squared:  0.0474
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 95 weights are ~= 1. The remaining 1053 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.393  0.873  0.948  0.908  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.71e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.319 1          1.523
## LastAuthorFemale 2.320 1          1.523
## Year      1.169 16          1.005

```

## Residuals from first and last author



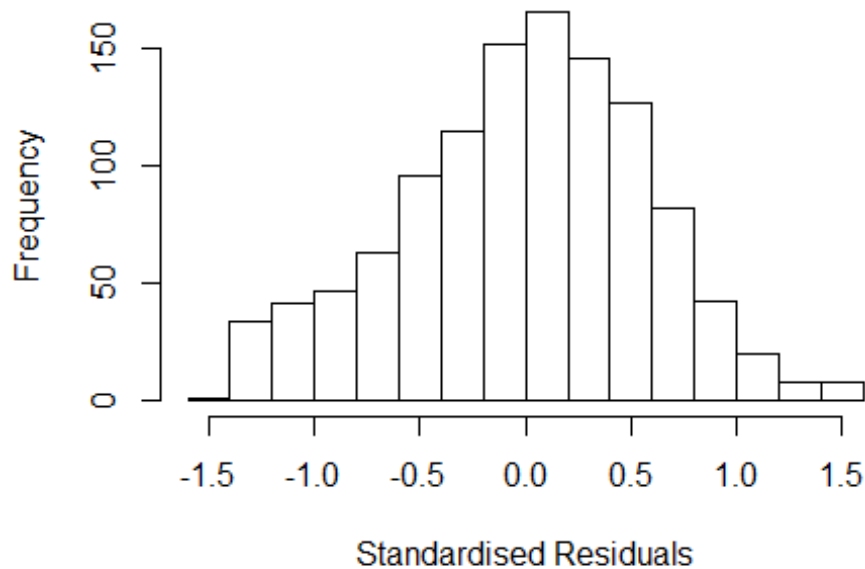
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3726 -0.3802  0.0297  0.3997  1.5856
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.17653    0.12946   9.09  <2e-16 ***
## FirstAuthorFemale1  0.02383    0.05323   0.45   0.654
## LastAuthorFemale1  0.05688    0.05360   1.06   0.289
## Year1997        -0.33830    0.16331  -2.07   0.039 *
## Year1998        -0.00464    0.15540  -0.03   0.976
## Year1999         0.06902    0.15057   0.46   0.647
## Year2000         0.19608    0.16211   1.21   0.227
## Year2001        -0.04932    0.17727  -0.28   0.781
## Year2002        -0.09421    0.14753  -0.64   0.523
## Year2003        -0.07751    0.15062  -0.51   0.607
## Year2004         0.05456    0.15426   0.35   0.724
## Year2005         0.07910    0.14353   0.55   0.582
```

```

## Year2006          -0.13913      0.14452      -0.96      0.336
## Year2007          -0.04101      0.15087      -0.27      0.786
## Year2008           0.02279      0.14822       0.15      0.878
## Year2009           0.01869      0.14014       0.13      0.894
## Year2010          -0.02424      0.14229      -0.17      0.865
## Year2011           0.16510      0.13991       1.18      0.238
## Year2012           0.10537      0.14419       0.73      0.465
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.576
## Multiple R-squared:  0.0381, Adjusted R-squared:  0.0228
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 1061 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.428  0.871  0.948  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.091 1      1.045
## Year              1.091 16      1.003

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4188 -0.3824 0.0293 0.3965 1.5823
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18059 0.12914 9.14 <2e-16 ***
## FirstAuthorFemale1 0.06582 0.03655 1.80 0.072 .
## Year1997 -0.34211 0.16347 -2.09 0.037 *
## Year1998 0.00166 0.15524 0.01 0.991
## Year1999 0.06592 0.15045 0.44 0.661
## Year2000 0.19008 0.16200 1.17 0.241
## Year2001 -0.05056 0.17619 -0.29 0.774
## Year2002 -0.08668 0.14714 -0.59 0.556
## Year2003 -0.07653 0.15033 -0.51 0.611
## Year2004 0.05676 0.15363 0.37 0.712
## Year2005 0.07963 0.14335 0.56 0.579
## Year2006 -0.13991 0.14435 -0.97 0.333
```

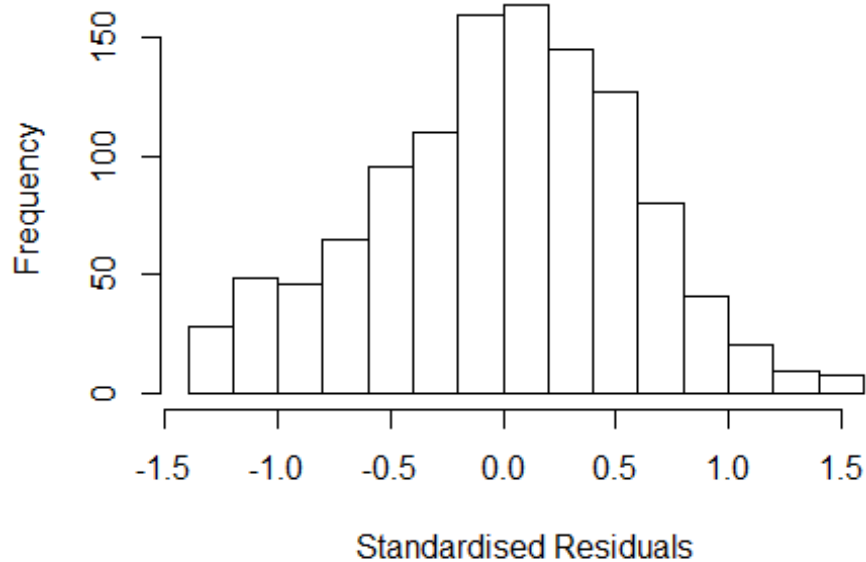


```

## Year2007      -0.03757    0.15090   -0.25    0.803
## Year2008      0.02665    0.14807    0.18    0.857
## Year2009      0.02094    0.13993    0.15    0.881
## Year2010     -0.01670    0.14195   -0.12    0.906
## Year2011      0.17236    0.13940    1.24    0.217
## Year2012      0.10883    0.14399    0.76    0.450
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.575
## Multiple R-squared:  0.0371, Adjusted R-squared:  0.0227
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 89 weights are ~= 1. The remaining 1059 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.430  0.873  0.948  0.906  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.092 1      1.045
## Year      1.092 16      1.003

```

## Residuals from last author



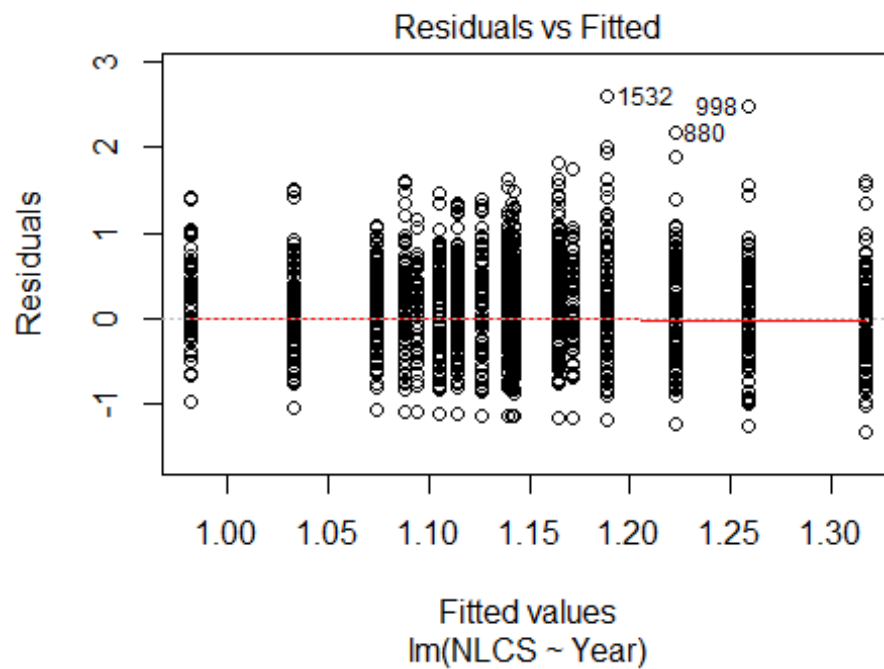
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3777 -0.3799 0.0293 0.3964 1.5814
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17916 0.12984 9.08 <2e-16 ***
## LastAuthorFemale1 0.07486 0.03678 2.04 0.042 *
## Year1997 -0.33795 0.16365 -2.07 0.039 *
## Year1998 -0.00586 0.15548 -0.04 0.970
## Year1999 0.07003 0.15110 0.46 0.643
## Year2000 0.19853 0.16195 1.23 0.221
## Year2001 -0.04910 0.17744 -0.28 0.782
## Year2002 -0.09545 0.14794 -0.65 0.519
## Year2003 -0.07758 0.15101 -0.51 0.608
## Year2004 0.05480 0.15478 0.35 0.723
## Year2005 0.08018 0.14378 0.56 0.577
## Year2006 -0.13760 0.14471 -0.95 0.342
```

```

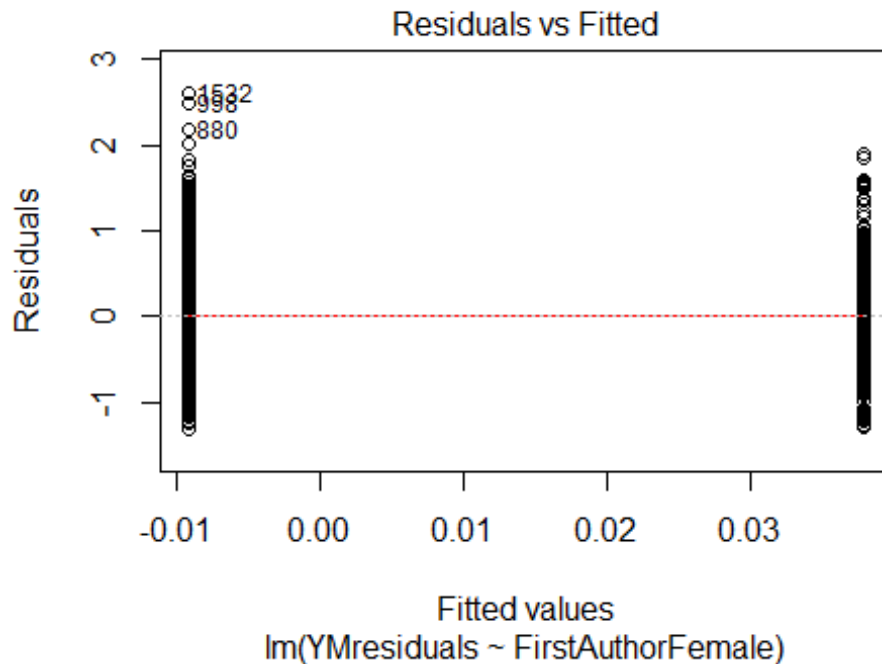
## Year2007          -0.04012      0.15107    -0.27      0.791
## Year2008           0.02245      0.14856      0.15      0.880
## Year2009           0.01894      0.14048      0.13      0.893
## Year2010          -0.02575      0.14272    -0.18      0.857
## Year2011           0.16540      0.14028      1.18      0.239
## Year2012           0.10658      0.14434      0.74      0.460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.575
## Multiple R-squared:  0.038, Adjusted R-squared:  0.0235
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 1062 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.429  0.872  0.949  0.906  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.71e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1148"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3318"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   95  103   89  117  147  149  131   90  113   96  123  143  166  189  204
## 2011 2012
##  251  245
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   89   97   72   99  123  110  114   75   99   85  101  115  142  170  180
## 2011 2012

```

```
## 221 201
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 89 96 72 98 121 107 111 72 94 85 100 112 134 166 173
## 2011 2012
## 208 196
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 29, df = 16, p-value = 0.03
```

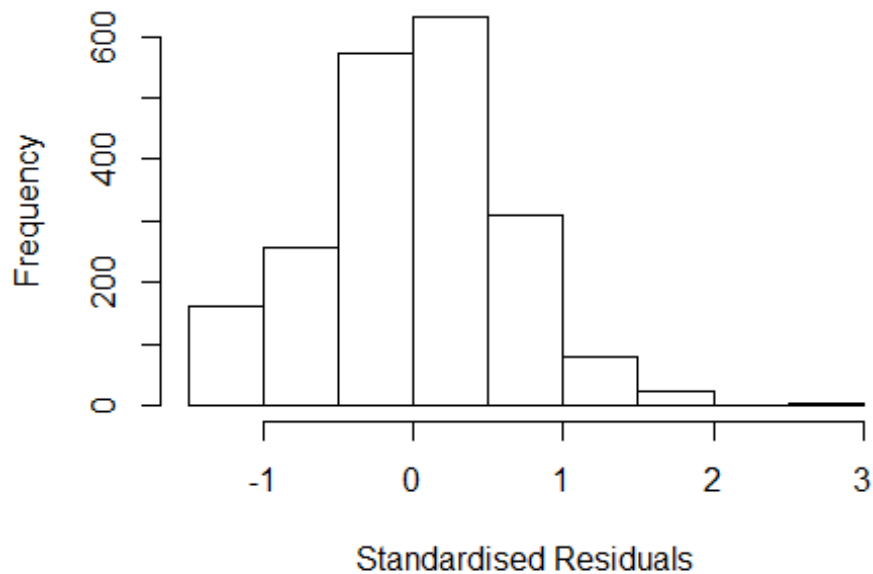


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.044, df = 1, p-value = 0.8
```



```
## [1] "Female first author team size 2018 geometric mean: 1.33488267230624"
## [1] "Male first author team size 2018 geometric mean: 1.18819207973285"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6300, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.30649922425326"
## [1] "Male last author team size 2018 geometric mean: 1.26207357243399"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6100, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.334 1      1.528
## LastAuthorFemale 2.293 1      1.514
## UniqueAuthors    1.287 4      1.032
## Year             1.272 16      1.008
```

## Residuals from first and last author and team size



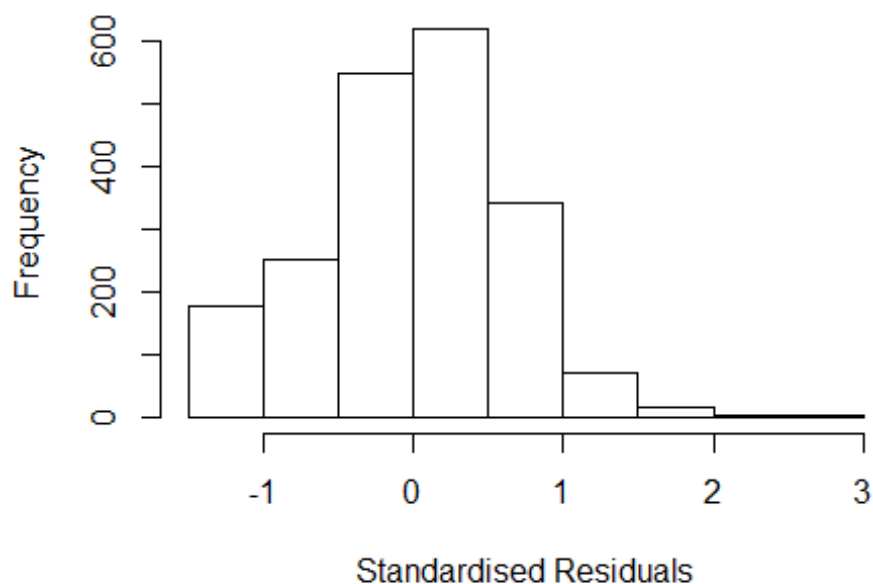
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 998   5044246868 3.743 2004      1213      3      2.586
## 1532 57749139264 3.783 2008      1213      3      2.694
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4179 -0.3972  0.0205  0.4142  2.6945
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9439    0.0708   13.34 < 2e-16 ***
## FirstAuthorFemale1 -0.0574    0.0524  -1.10  0.27355
## LastAuthorFemale1  0.0276    0.0520   0.53  0.59506
## UniqueAuthors2    0.3294    0.0372   8.85 < 2e-16 ***
## UniqueAuthors3    0.2854    0.0535   5.34  1.1e-07 ***
## UniqueAuthors4    0.3467    0.1024   3.39  0.00072 ***
## UniqueAuthors5    0.2654    0.2165   1.23  0.22031
## Year1997          0.1324    0.0948   1.40  0.16260
## Year1998          0.0878    0.0916   0.96  0.33828
```

```

## Year1999      0.0673      0.0865      0.78  0.43678
## Year2000      0.1186      0.0816      1.45  0.14630
## Year2001      0.1732      0.0868      2.00  0.04617 *
## Year2002      0.2360      0.0881      2.68  0.00744 **
## Year2003      0.1752      0.0925      1.89  0.05846 .
## Year2004      0.2429      0.0887      2.74  0.00622 **
## Year2005      0.3296      0.0900      3.66  0.00026 ***
## Year2006      0.1266      0.0894      1.42  0.15665
## Year2007      0.0768      0.0874      0.88  0.37926
## Year2008      0.1743      0.0898      1.94  0.05247 .
## Year2009      0.1198      0.0780      1.54  0.12479
## Year2010      0.1302      0.0805      1.62  0.10595
## Year2011      0.1225      0.0802      1.53  0.12660
## Year2012      0.1532      0.0829      1.85  0.06484 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.615
## Multiple R-squared:  0.0521, Adjusted R-squared:  0.0417
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 157 weights are ~= 1. The remaining 1877 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0155 0.8740 0.9520 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi      bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max      maxit.scale
##      500      50      2      1      1000      200
## trace.lev      mts      compute.rd
##      0      1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.912 1      1.707
## LastAuthorFemale 2.870 1      1.694
## Year      1.081 16      1.002

```

## Residuals from first and last author



```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 998   5044246868 3.743 2004     1213      3      2.525
## 1532 57749139264 3.783 2008     1213      3      2.641
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2893 -0.4138  0.0253  0.4292  2.6406
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0128    0.0710   14.26 < 2e-16 ***
## FirstAuthorFemale1 -0.0168    0.0616   -0.27  0.78496
## LastAuthorFemale1 -0.0347    0.0597   -0.58  0.56129
## Year1997          0.1496    0.0952    1.57  0.11604
## Year1998          0.1334    0.0948    1.41  0.15955
## Year1999          0.0477    0.0876    0.54  0.58631
## Year2000          0.1097    0.0821    1.34  0.18183
## Year2001          0.1756    0.0883    1.99  0.04700 *
## Year2002          0.2225    0.0902    2.47  0.01376 *
## Year2003          0.1851    0.0985    1.88  0.06036 .
## Year2004          0.2568    0.0913    2.81  0.00495 **
```

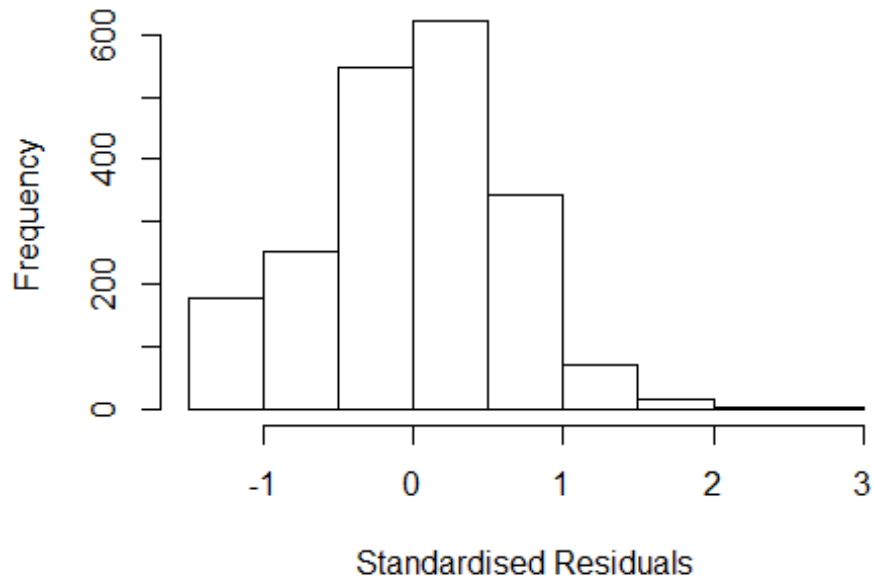


```

## Year2005          0.3280      0.0919      3.57  0.00037 ***
## Year2006          0.1331      0.0907      1.47  0.14226
## Year2007          0.1055      0.0867      1.22  0.22352
## Year2008          0.1812      0.0912      1.99  0.04714 *
## Year2009          0.1402      0.0790      1.77  0.07624 .
## Year2010          0.1607      0.0821      1.96  0.05027 .
## Year2011          0.1713      0.0801      2.14  0.03258 *
## Year2012          0.1837      0.0838      2.19  0.02853 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00259
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 177 weights are ~= 1. The remaining 1857 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0395 0.8720 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.055 1      1.027
## Year      1.055 16      1.002

```

## Residuals from first author



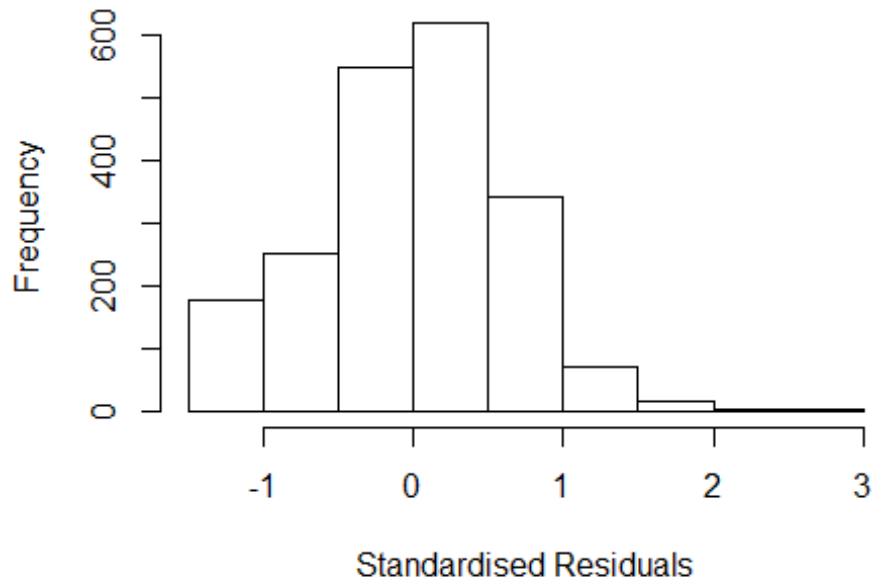
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 998   5044246868 3.743 2004     1213      3      2.525
## 1532 57749139264 3.783 2008     1213      3      2.641
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2913 -0.4139  0.0265  0.4277  2.6389
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0097    0.0708   14.26 < 2e-16 ***
## FirstAuthorFemale1 -0.0459    0.0370   -1.24  0.21488
## Year1997          0.1468    0.0949    1.55  0.12208
## Year1998          0.1342    0.0948    1.42  0.15707
## Year1999          0.0469    0.0876    0.54  0.59245
## Year2000          0.1085    0.0821    1.32  0.18688
## Year2001          0.1731    0.0882    1.96  0.04970 *
## Year2002          0.2214    0.0903    2.45  0.01431 *
## Year2003          0.1838    0.0986    1.87  0.06231 .
## Year2004          0.2548    0.0912    2.79  0.00526 **
## Year2005          0.3274    0.0921    3.56  0.00039 ***
```

```

## Year2006          0.1324      0.0908      1.46  0.14483
## Year2007          0.1033      0.0867      1.19  0.23332
## Year2008          0.1802      0.0912      1.98  0.04832 *
## Year2009          0.1401      0.0791      1.77  0.07658 .
## Year2010          0.1602      0.0822      1.95  0.05128 .
## Year2011          0.1707      0.0802      2.13  0.03334 *
## Year2012          0.1822      0.0838      2.17  0.02989 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.629
## Multiple R-squared:  0.0113, Adjusted R-squared:  0.00293
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 177 weights are ~= 1. The remaining 1857 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0396 0.8730 0.9510 0.9080 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.041 1      1.020
## Year      1.041 16      1.001

```

## Residuals from last author



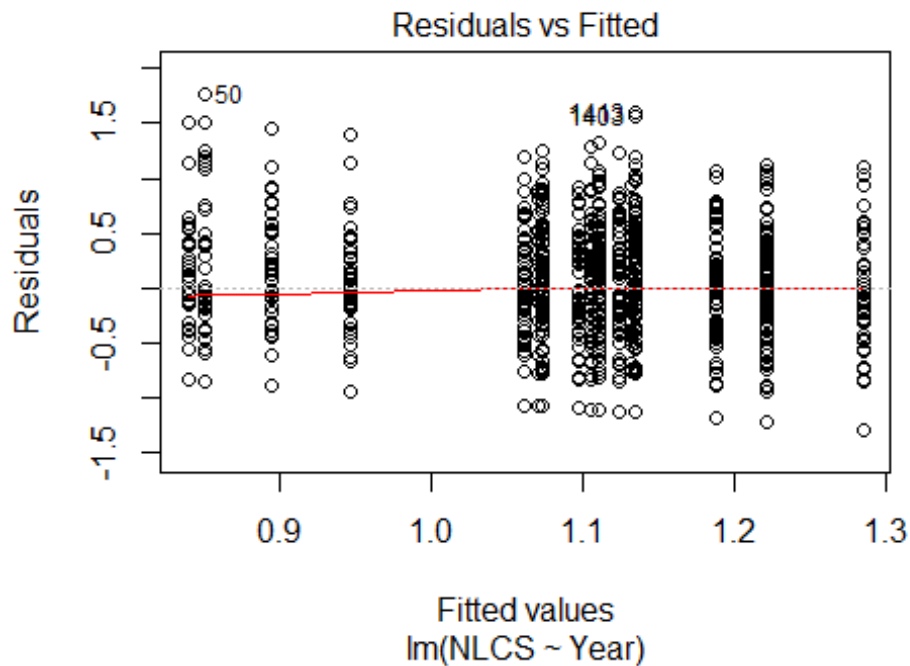
```
## [1] "List of 2 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 998   5044246868 3.743 2004     1213      3      2.525
## 1532 57749139264 3.783 2008     1213      3      2.641
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2898 -0.4143  0.0257  0.4300  2.6401
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0094     0.0697   14.48 < 2e-16 ***
## LastAuthorFemale1 -0.0481     0.0359   -1.34  0.18100
## Year1997          0.1507     0.0950    1.59  0.11305
## Year1998          0.1328     0.0947    1.40  0.16124
## Year1999          0.0479     0.0875    0.55  0.58434
## Year2000          0.1096     0.0821    1.34  0.18182
## Year2001          0.1763     0.0883    2.00  0.04608 *
## Year2002          0.2228     0.0902    2.47  0.01361 *
## Year2003          0.1855     0.0986    1.88  0.06004 .
## Year2004          0.2578     0.0913    2.82  0.00479 **
## Year2005          0.3285     0.0918    3.58  0.00035 ***
```

```

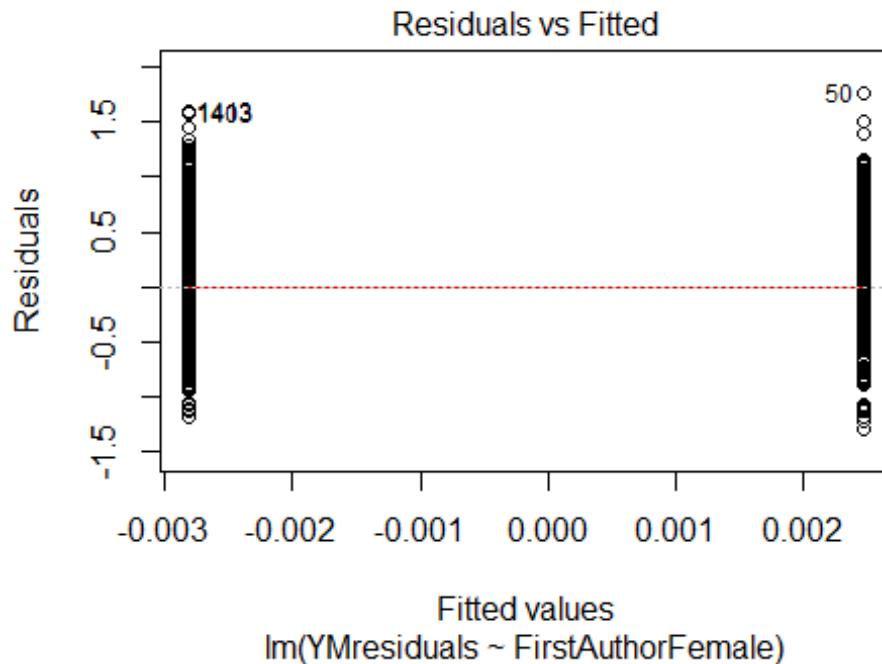
## Year2006          0.1336      0.0907      1.47  0.14078
## Year2007          0.1063      0.0866      1.23  0.22001
## Year2008          0.1816      0.0912      1.99  0.04663 *
## Year2009          0.1404      0.0790      1.78  0.07559 .
## Year2010          0.1615      0.0818      1.97  0.04868 *
## Year2011          0.1716      0.0801      2.14  0.03236 *
## Year2012          0.1844      0.0838      2.20  0.02789 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.63
## Multiple R-squared:  0.0114, Adjusted R-squared:  0.00305
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 178 weights are ~= 1. The remaining 1856 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0398 0.8730 0.9510 0.9090 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.92e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2034"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3319"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   61   52   70   67   78   72   57   31   50   44   44   78   74   97  121
## 2011 2012
##  139  129
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   38   49   47   61   59   48   30   47   38   35   71   67   83  106

```

```
## 2011 2012
## 124 116
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 50 36 48 46 59 56 47 27 45 38 33 67 62 75 101
## 2011 2012
## 113 107
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.09
```

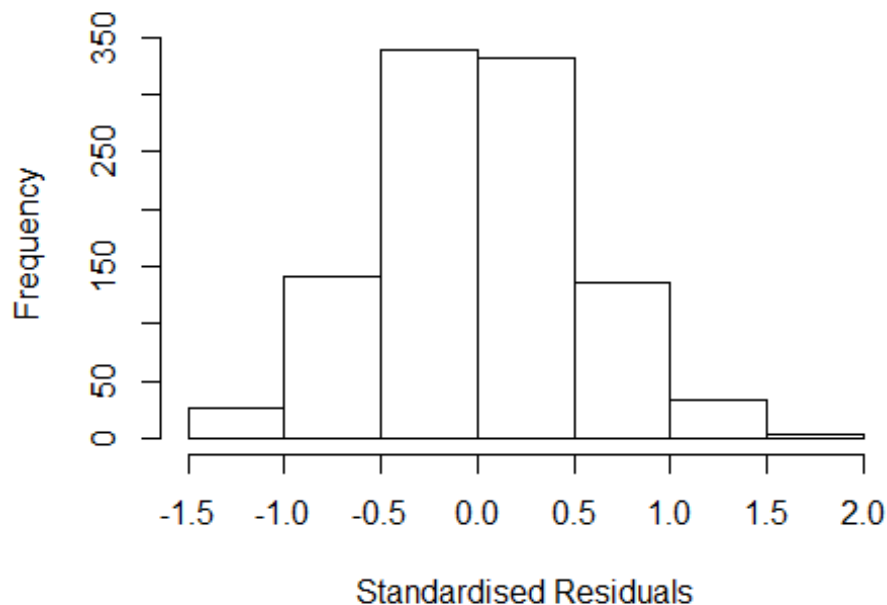


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.84, df = 1, p-value = 0.4
```



```
## [1] "Female first author team size 2018 geometric mean: 2.00114910388434"
## [1] "Male first author team size 2018 geometric mean: 1.80476129689686"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2000, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.91863411713842"
## [1] "Male last author team size 2018 geometric mean: 1.92402825747101"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1900, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.714 1      1.309
## LastAuthorFemale  1.648 1      1.284
## UniqueAuthors     1.340 4      1.037
## Year               1.534 16     1.013
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.36109 -0.35956 -0.00175  0.35896  1.80229
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.6778    0.1042    6.51 1.2e-10 ***
## FirstAuthorFemale1  0.0066    0.0440    0.15 0.88071
## LastAuthorFemale1 -0.0164    0.0430   -0.38 0.70371
## UniqueAuthors2    0.1263    0.0416    3.03 0.00248 **
## UniqueAuthors3    0.1076    0.0523    2.06 0.03999 *
## UniqueAuthors4    0.3079    0.0736    4.18 3.1e-05 ***
## UniqueAuthors5    0.2919    0.0983    2.97 0.00305 **
## Year1997          0.0860    0.1384    0.62 0.53448
## Year1998          0.1260    0.1338    0.94 0.34636
## Year1999          0.1913    0.1307    1.46 0.14372
```

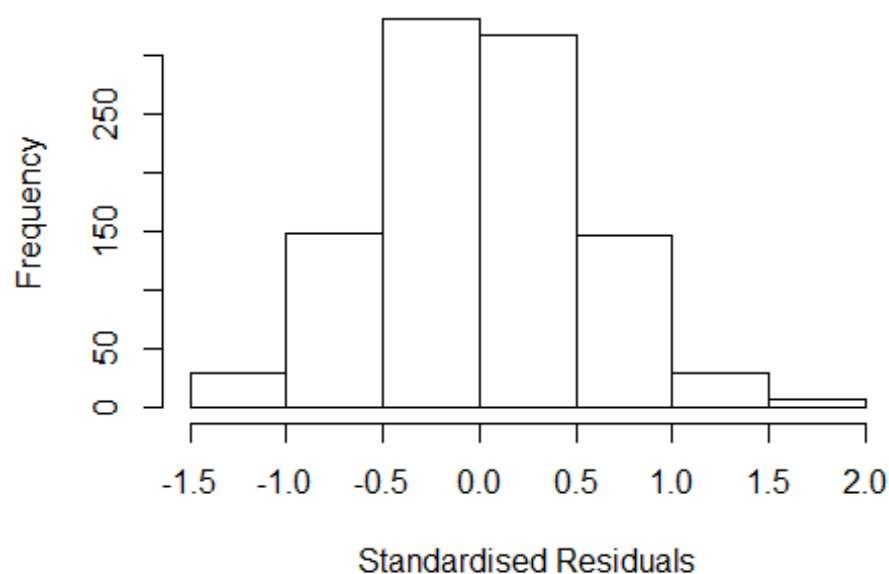


```

## Year2000          0.3702      0.1268      2.92  0.00358 **
## Year2001          0.3705      0.1260      2.94  0.00335 **
## Year2002          0.3556      0.1278      2.78  0.00548 **
## Year2003          0.3546      0.1415      2.51  0.01237 *
## Year2004          0.5192      0.1317      3.94  8.6e-05 ***
## Year2005          0.4426      0.1262      3.51  0.00047 ***
## Year2006          0.3661      0.1278      2.86  0.00427 **
## Year2007          0.3038      0.1207      2.52  0.01196 *
## Year2008          0.4718      0.1232      3.83  0.00014 ***
## Year2009          0.4474      0.1168      3.83  0.00014 ***
## Year2010          0.3015      0.1196      2.52  0.01183 *
## Year2011          0.3306      0.1200      2.75  0.00599 **
## Year2012          0.3687      0.1237      2.98  0.00295 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0842, Adjusted R-squared:  0.0637
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 86 weights are ~= 1. The remaining 924 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  0.871  0.951  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.90e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.660 1          1.288
## LastAuthorFemale  1.576 1          1.256
## Year              1.168 16          1.005

```

## Residuals from first and last author



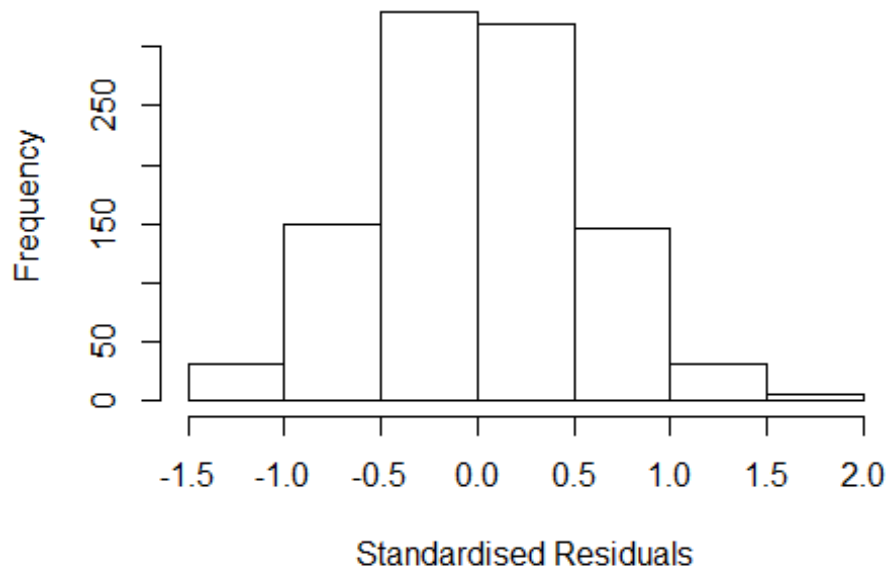
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25538 -0.36243 -0.00842 0.36400 1.87035
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7197 0.1041 6.91 8.5e-12 ***
## FirstAuthorFemale1 0.0230 0.0444 0.52 0.60540
## LastAuthorFemale1 -0.0322 0.0433 -0.74 0.45739
## Year1997 0.0849 0.1351 0.63 0.52992
## Year1998 0.1546 0.1360 1.14 0.25602
## Year1999 0.2046 0.1278 1.60 0.10980
## Year2000 0.4177 0.1254 3.33 0.00089 ***
## Year2001 0.3969 0.1264 3.14 0.00174 **
## Year2002 0.4007 0.1287 3.11 0.00190 **
## Year2003 0.3800 0.1476 2.57 0.01020 *
## Year2004 0.5449 0.1343 4.06 5.4e-05 ***
## Year2005 0.5184 0.1274 4.07 5.1e-05 ***
```

```

## Year2006          0.4000      0.1317      3.04  0.00244 **
## Year2007          0.3541      0.1195      2.96  0.00311 **
## Year2008          0.4919      0.1232      3.99  7.0e-05 ***
## Year2009          0.4877      0.1160      4.20  2.9e-05 ***
## Year2010          0.3455      0.1171      2.95  0.00324 **
## Year2011          0.3673      0.1193      3.08  0.00214 **
## Year2012          0.4072      0.1232      3.30  0.00099 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.0565, Adjusted R-squared:  0.0394
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 80 weights are ~= 1. The remaining 930 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.209  0.867  0.951  0.909  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      9.90e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.132 1          1.064
## Year              1.132 16          1.004

```

## Residuals from first author



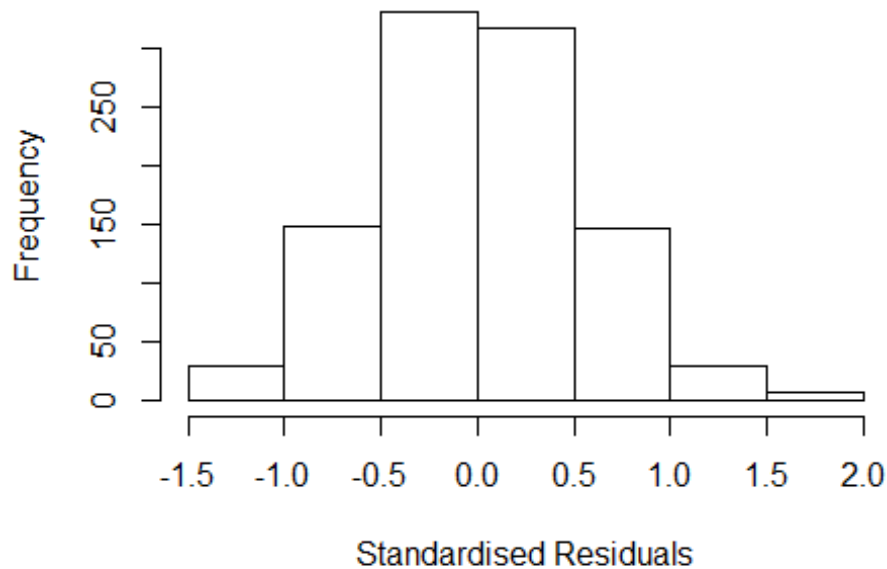
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26154 -0.36289 -0.00807 0.36206 1.89518
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.71299 0.10320 6.91 8.7e-12 ***
## FirstAuthorFemale1 0.00484 0.03671 0.13 0.89518
## Year1997 0.08774 0.13497 0.65 0.51582
## Year1998 0.15605 0.13572 1.15 0.25050
## Year1999 0.20677 0.12785 1.62 0.10613
## Year2000 0.41709 0.12508 3.33 0.00089 ***
## Year2001 0.39867 0.12657 3.15 0.00168 **
## Year2002 0.39965 0.12848 3.11 0.00192 **
## Year2003 0.38275 0.14763 2.59 0.00967 **
## Year2004 0.54371 0.13467 4.04 5.8e-05 ***
## Year2005 0.52087 0.12715 4.10 4.5e-05 ***
## Year2006 0.39682 0.13167 3.01 0.00265 **
```

```

## Year2007          0.35316    0.11935    2.96  0.00316 **
## Year2008          0.49139    0.12317    3.99  7.1e-05 ***
## Year2009          0.48870    0.11572    4.22  2.6e-05 ***
## Year2010          0.34161    0.11724    2.91  0.00365 **
## Year2011          0.36872    0.11921    3.09  0.00204 **
## Year2012          0.40708    0.12329    3.30  0.00099 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.056, Adjusted R-squared:  0.0399
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 934 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.195  0.865   0.950   0.909   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.90e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.075 1          1.037
## Year              1.075 16          1.002

```

## Residuals from last author



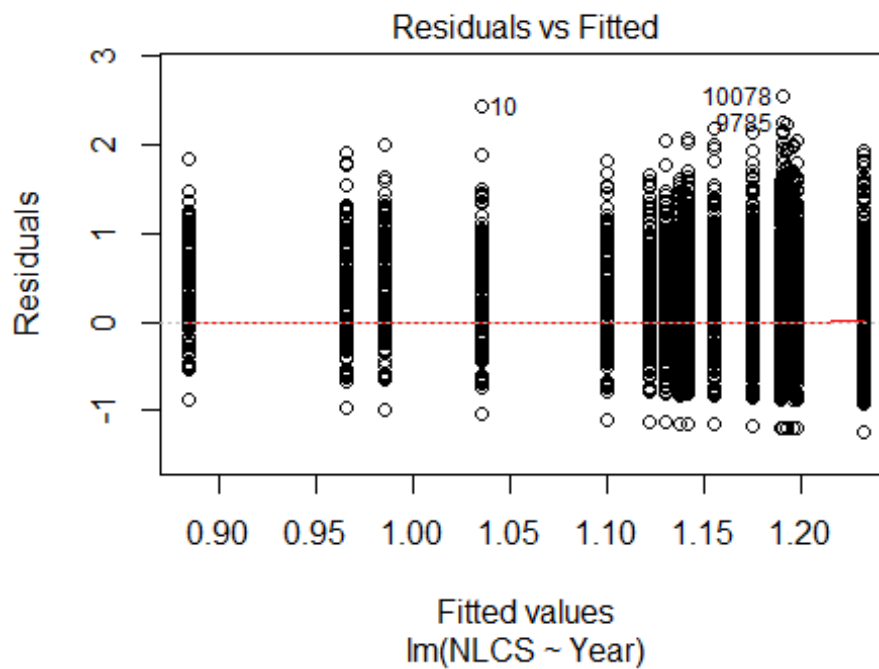
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.25140 -0.36172 -0.00911 0.36687 1.88979
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7232 0.1046 6.92 8.3e-12 ***
## LastAuthorFemale1 -0.0194 0.0358 -0.54 0.58767
## Year1997 0.0861 0.1352 0.64 0.52467
## Year1998 0.1558 0.1358 1.15 0.25167
## Year1999 0.2052 0.1278 1.61 0.10879
## Year2000 0.4170 0.1254 3.33 0.00091 ***
## Year2001 0.3982 0.1264 3.15 0.00169 **
## Year2002 0.4021 0.1286 3.13 0.00181 **
## Year2003 0.3836 0.1474 2.60 0.00939 **
## Year2004 0.5476 0.1340 4.09 4.8e-05 ***
## Year2005 0.5211 0.1272 4.10 4.5e-05 ***
## Year2006 0.4037 0.1309 3.08 0.00209 **
```

```

## Year2007          0.3557      0.1195      2.98  0.00298 **
## Year2008          0.4950      0.1230      4.03  6.1e-05 ***
## Year2009          0.4901      0.1156      4.24  2.5e-05 ***
## Year2010          0.3470      0.1170      2.96  0.00310 **
## Year2011          0.3707      0.1189      3.12  0.00187 **
## Year2012          0.4127      0.1217      3.39  0.00072 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.0564, Adjusted R-squared:  0.0402
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 85 weights are ~= 1. The remaining 925 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.197  0.867  0.950  0.908  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.90e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1010"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3320"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 241 272 291 296 384 422 446 407 484 509 580 685 690 896 1074
## 2011 2012
## 1175 1221
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 207 230 233 220 286 241 363 345 420 452 506 586 579 765 932
## 2011 2012

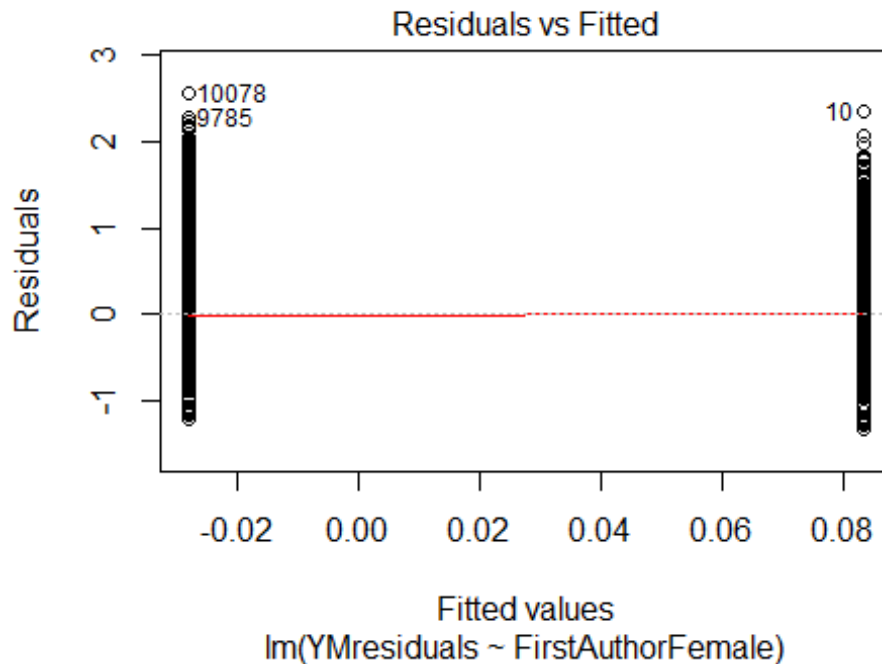
```

```
## 993 1055
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 198 229 230 218 281 235 359 341 409 449 499 573 561 758 904
## 2011 2012
## 973 1040
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 42, df = 16, p-value = 4e-04
```



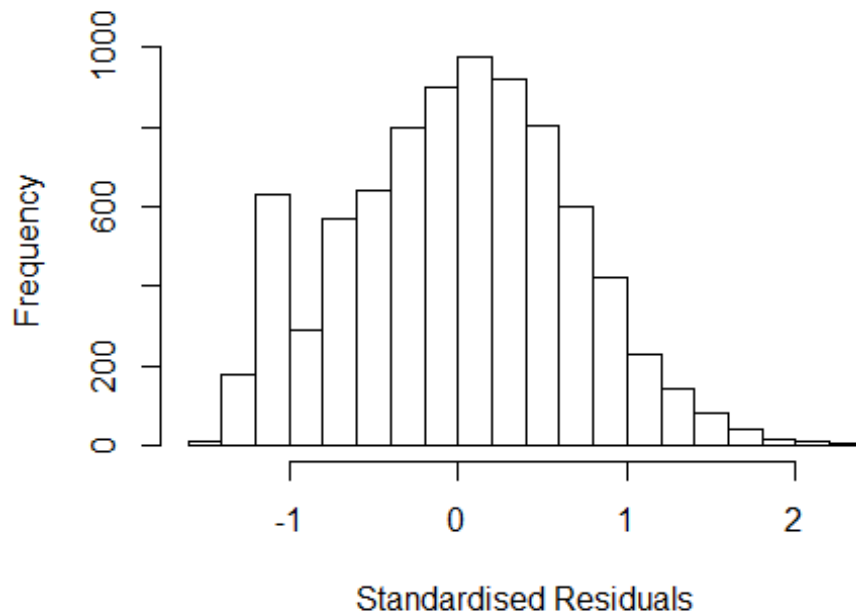
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.7, df = 1, p-value = 0.03
```





```
## [1] "Female first author team size 2018 geometric mean: 1.18966891641118"
## [1] "Male first author team size 2018 geometric mean: 1.18641567096404"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 99000, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.21069222490476"
## [1] "Male last author team size 2018 geometric mean: 1.17576279775307"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1e+05, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.584 1      1.893
## LastAuthorFemale  3.581 1      1.892
## UniqueAuthors    1.049 4      1.006
## Year             1.043 16     1.001
```

## Residuals from first and last author and team size



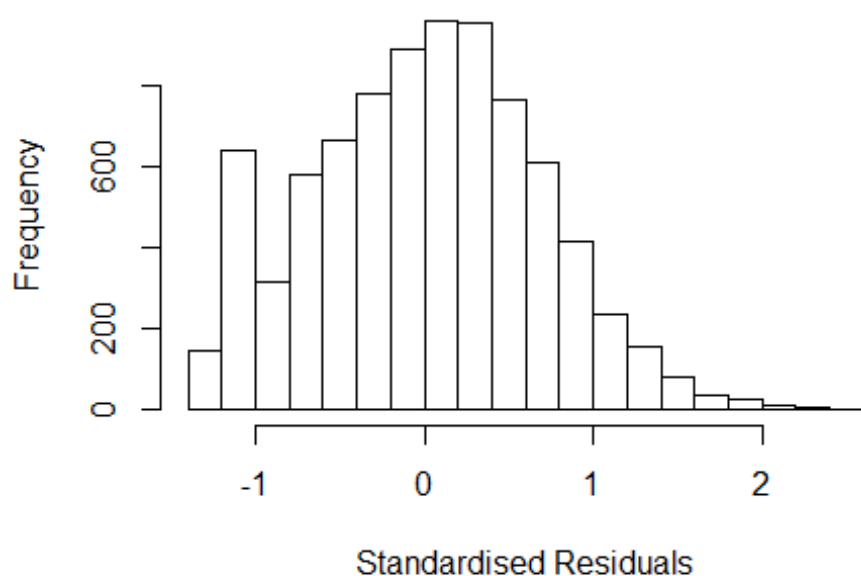
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.5214 -0.4605 0.0233 0.4601 2.3821
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9533 0.0558 17.08 < 2e-16 ***
## FirstAuthorFemale1 0.0550 0.0328 1.67 0.09407 .
## LastAuthorFemale1 0.0766 0.0327 2.34 0.01927 *
## UniqueAuthors2 0.1983 0.0238 8.32 < 2e-16 ***
## UniqueAuthors3 0.2709 0.0453 5.98 2.3e-09 ***
## UniqueAuthors4 0.1774 0.0836 2.12 0.03376 *
## UniqueAuthors5 0.5094 0.1381 3.69 0.00023 ***
## Year1997 0.0698 0.0765 0.91 0.36200
## Year1998 -0.0688 0.0733 -0.94 0.34804
## Year1999 -0.1402 0.0739 -1.90 0.05766 .
```

```

## Year2000          0.0873      0.0698      1.25  0.21132
## Year2001         -0.0448      0.0771     -0.58  0.56067
## Year2002          0.1301      0.0690      1.88  0.05963 .
## Year2003          0.1130      0.0669      1.69  0.09124 .
## Year2004          0.1765      0.0653      2.70  0.00691 **
## Year2005          0.1693      0.0629      2.69  0.00715 **
## Year2006          0.1129      0.0624      1.81  0.07020 .
## Year2007          0.2062      0.0627      3.29  0.00102 **
## Year2008          0.1486      0.0625      2.38  0.01735 *
## Year2009          0.1200      0.0604      1.98  0.04721 *
## Year2010          0.1707      0.0600      2.85  0.00444 **
## Year2011          0.1656      0.0600      2.76  0.00583 **
## Year2012          0.1574      0.0602      2.62  0.00892 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.69
## Multiple R-squared:  0.0322, Adjusted R-squared:  0.0296
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 672 weights are ~= 1. The remaining 7585 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.872  0.951  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.879 1      1.970
## LastAuthorFemale 3.876 1      1.969
## Year              1.013 16      1.000

```

## Residuals from first and last author

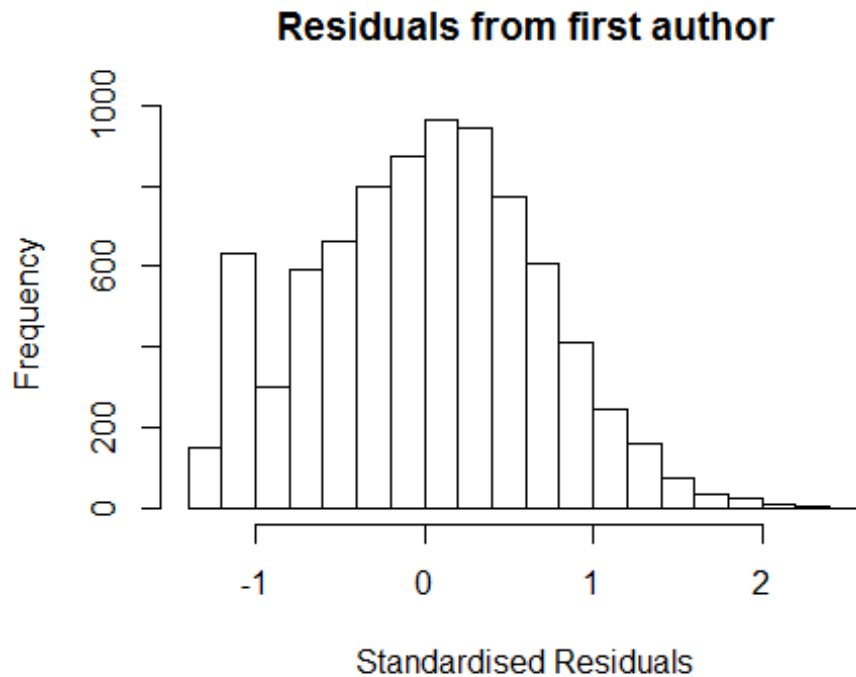


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.329 -0.470  0.027  0.470  2.484
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9835    0.0559   17.59 < 2e-16 ***
## FirstAuthorFemale1 0.0350    0.0345    1.02  0.30905
## LastAuthorFemale1 0.0997    0.0342    2.91  0.00357 **
## Year1997         0.0747    0.0772    0.97  0.33302
## Year1998        -0.0657    0.0736   -0.89  0.37231
## Year1999        -0.1440    0.0748   -1.93  0.05425 .
## Year2000         0.0927    0.0700    1.33  0.18513
## Year2001        -0.0414    0.0774   -0.53  0.59330
## Year2002         0.1243    0.0695    1.79  0.07378 .
## Year2003         0.1066    0.0671    1.59  0.11194
## Year2004         0.1724    0.0658    2.62  0.00883 **
## Year2005         0.1742    0.0632    2.75  0.00589 **
```

```

## Year2006          0.1173      0.0624      1.88  0.06045 .
## Year2007          0.2105      0.0629      3.35  0.00082 ***
## Year2008          0.1520      0.0629      2.42  0.01568 *
## Year2009          0.1216      0.0606      2.00  0.04504 *
## Year2010          0.1776      0.0602      2.95  0.00320 **
## Year2011          0.1682      0.0602      2.79  0.00522 **
## Year2012          0.1554      0.0604      2.57  0.01006 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.696
## Multiple R-squared:  0.0192, Adjusted R-squared:  0.0171
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 706 weights are ~= 1. The remaining 7551 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.176  0.872  0.950  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.007 1      1.004
## Year              1.007 16      1.000

```



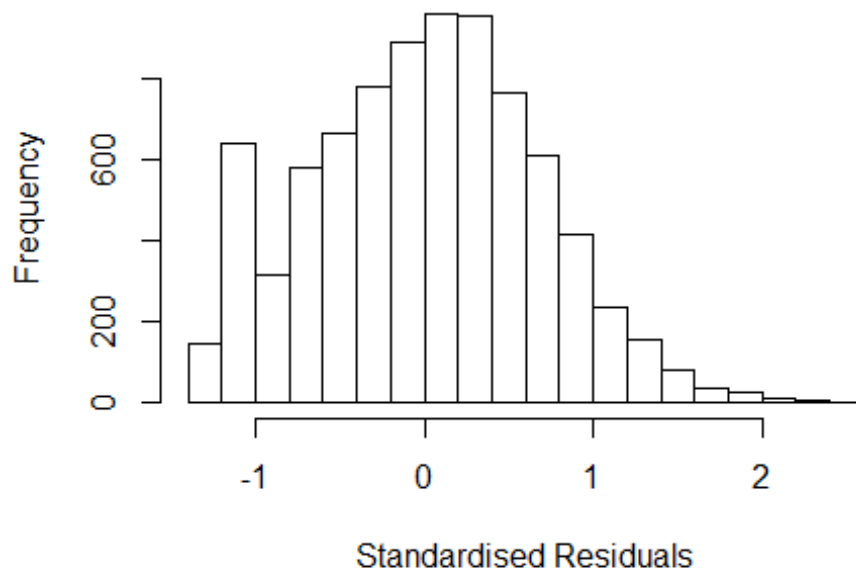
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.320 -0.471  0.026  0.470  2.579
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9862    0.0559   17.63 < 2e-16 ***
## FirstAuthorFemale1 0.1213    0.0175    6.93 4.7e-12 ***
## Year1997        0.0775    0.0772    1.00 0.31542
## Year1998       -0.0620    0.0736   -0.84 0.40015
## Year1999       -0.1455    0.0747   -1.95 0.05130 .
## Year2000        0.0935    0.0700    1.34 0.18168
## Year2001       -0.0416    0.0775   -0.54 0.59130
## Year2002        0.1277    0.0694    1.84 0.06592 .
## Year2003        0.1090    0.0671    1.62 0.10453
## Year2004        0.1735    0.0659    2.63 0.00847 **
## Year2005        0.1747    0.0633    2.76 0.00577 **
## Year2006        0.1183    0.0625    1.89 0.05844 .
```

```

## Year2007          0.2120      0.0629      3.37  0.00076 ***
## Year2008          0.1535      0.0630      2.44  0.01482 *
## Year2009          0.1212      0.0607      2.00  0.04587 *
## Year2010          0.1787      0.0603      2.97  0.00302 **
## Year2011          0.1705      0.0602      2.83  0.00465 **
## Year2012          0.1577      0.0604      2.61  0.00903 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.697
## Multiple R-squared:  0.0182, Adjusted R-squared:  0.0161
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 718 weights are ~= 1. The remaining 7539 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.142  0.872  0.950  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.007 1      1.003
## Year      1.007 16      1.000

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3246 -0.4708  0.0268  0.4704  2.4533
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9847    0.0559   17.61 < 2e-16 ***
## LastAuthorFemale1 0.1295    0.0174    7.43 1.2e-13 ***
## Year1997        0.0740    0.0772    0.96 0.33814
## Year1998       -0.0671    0.0736   -0.91 0.36248
## Year1999       -0.1436    0.0748   -1.92 0.05503 .
## Year2000        0.0929    0.0700    1.33 0.18430
## Year2001       -0.0415    0.0774   -0.54 0.59171
## Year2002        0.1233    0.0696    1.77 0.07618 .
## Year2003        0.1059    0.0671    1.58 0.11452
## Year2004        0.1723    0.0658    2.62 0.00889 **
## Year2005        0.1743    0.0632    2.76 0.00586 **
## Year2006        0.1176    0.0625    1.88 0.05971 .
```

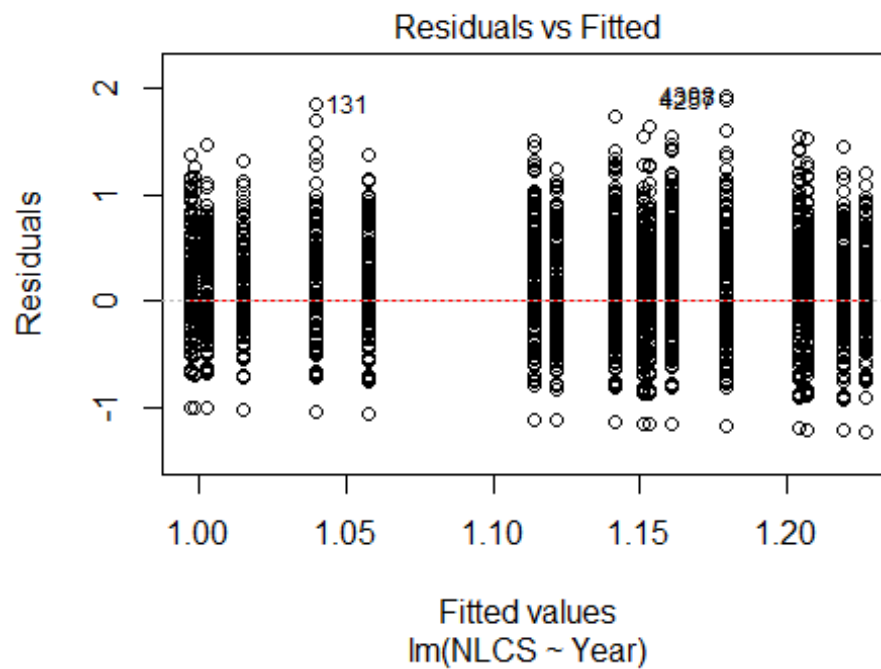


```

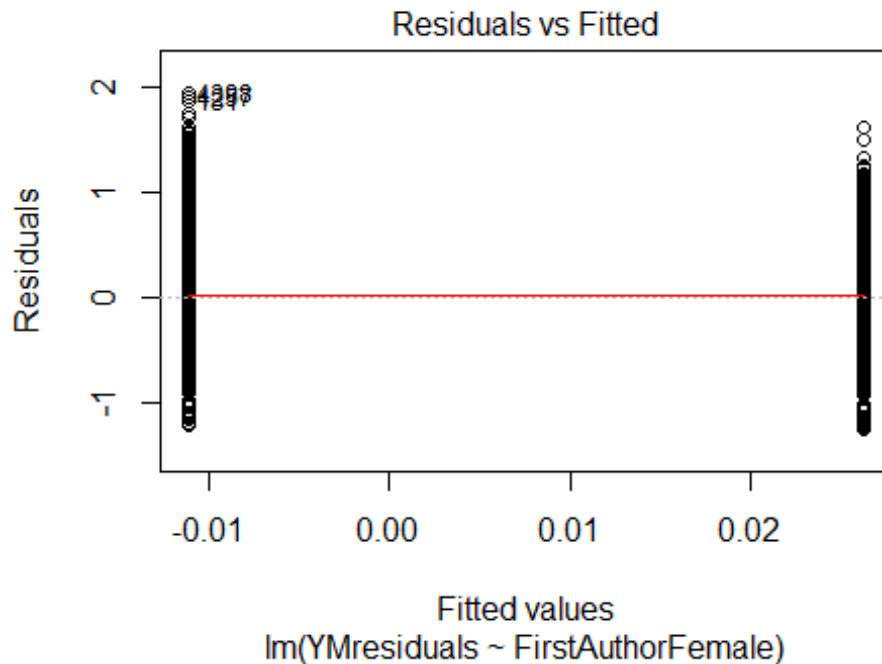
## Year2007          0.2104      0.0629      3.35  0.00082 ***
## Year2008          0.1518      0.0629      2.41  0.01585 *
## Year2009          0.1220      0.0606      2.01  0.04421 *
## Year2010          0.1775      0.0602      2.95  0.00321 **
## Year2011          0.1680      0.0602      2.79  0.00528 **
## Year2012          0.1555      0.0604      2.58  0.01000 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.696
## Multiple R-squared:  0.0191, Adjusted R-squared:  0.0171
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 704 weights are ~= 1. The remaining 7553 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.188  0.871  0.950  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.21e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 8257"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3321"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  202  169  189  204  230  223  233  229  208  235  241  284  249  304  294
## 2011 2012
##  339  344
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  175  150  150  157  154  118  212  208  188  209  215  257  219  270  261
## 2011 2012

```

```
## 305 301
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 169 147 146 155 150 114 203 200 181 201 210 248 212 264 248
## 2011 2012
## 297 290
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 39, df = 16, p-value = 0.001
```

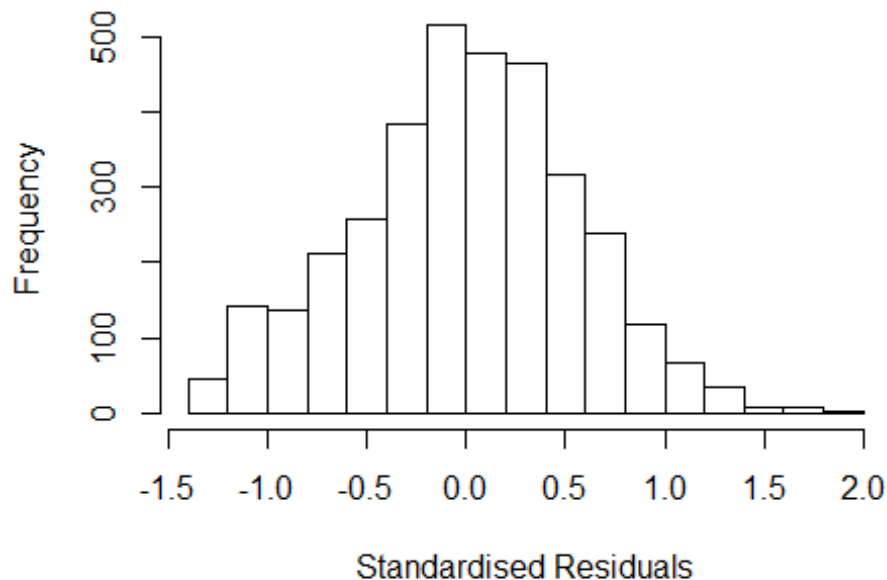


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.3, df = 1, p-value = 0.04
```



```
## [1] "Female first author team size 2018 geometric mean: 1.92402986135143"
## [1] "Male first author team size 2018 geometric mean: 1.50406239992707"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7900, p-value = 0.001
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.85064566307523"
## [1] "Male last author team size 2018 geometric mean: 1.56296113110559"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7300, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.492 1          1.222
## LastAuthorFemale  1.503 1          1.226
## UniqueAuthors    1.087 4          1.010
## Year              1.113 16         1.003
```

## Residuals from first and last author and team size



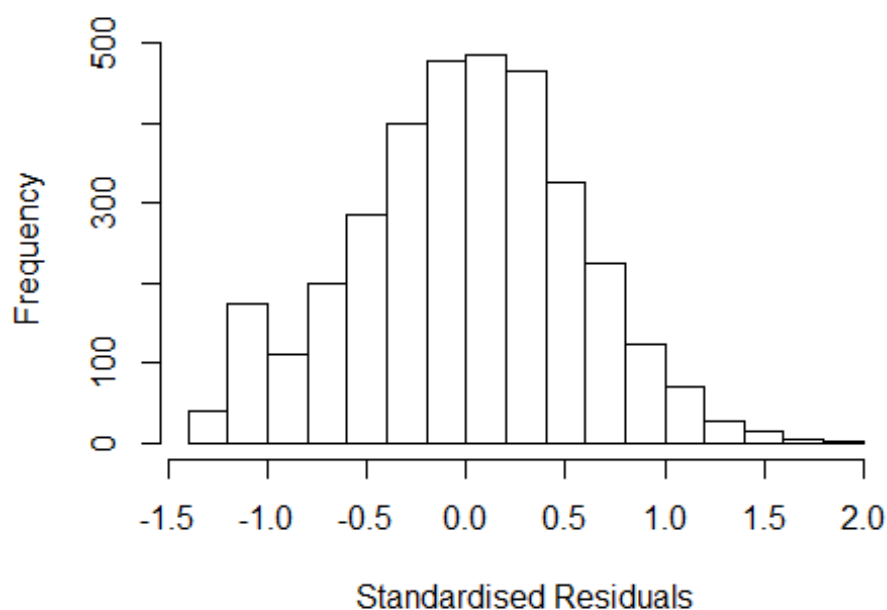
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.32698 -0.35705 0.00839 0.36972 1.97363
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.96801 0.04777 20.26 < 2e-16 ***
## FirstAuthorFemale1 0.02234 0.02588 0.86 0.3881
## LastAuthorFemale1 0.01833 0.02601 0.70 0.4809
## UniqueAuthors2 0.11881 0.02304 5.16 2.7e-07 ***
## UniqueAuthors3 0.14670 0.03134 4.68 3.0e-06 ***
## UniqueAuthors4 0.15274 0.05793 2.64 0.0084 **
## UniqueAuthors5 0.10087 0.07361 1.37 0.1707
## Year1997 -0.02506 0.06416 -0.39 0.6961
## Year1998 -0.03549 0.07046 -0.50 0.6144
## Year1999 0.02657 0.06707 0.40 0.6920
```

```

## Year2000      -0.00159    0.06697   -0.02    0.9811
## Year2001      -0.05131    0.06720   -0.76    0.4452
## Year2002       0.17160    0.06289    2.73    0.0064 **
## Year2003       0.18361    0.05925    3.10    0.0020 **
## Year2004       0.19833    0.06032    3.29    0.0010 **
## Year2005       0.10295    0.06239    1.65    0.0990 .
## Year2006       0.12379    0.05964    2.08    0.0380 *
## Year2007       0.12293    0.05738    2.14    0.0322 *
## Year2008       0.18261    0.06180    2.96    0.0031 **
## Year2009       0.11423    0.06138    1.86    0.0628 .
## Year2010       0.08538    0.06041    1.41    0.1577
## Year2011       0.12289    0.05971    2.06    0.0396 *
## Year2012       0.15836    0.06163    2.57    0.0102 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.0325, Adjusted R-squared:  0.0263
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 282 weights are ~= 1. The remaining 3153 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.176  0.867  0.952  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.510 1      1.229
## LastAuthorFemale  1.524 1      1.234
## Year              1.034 16      1.001

```

## Residuals from first and last author



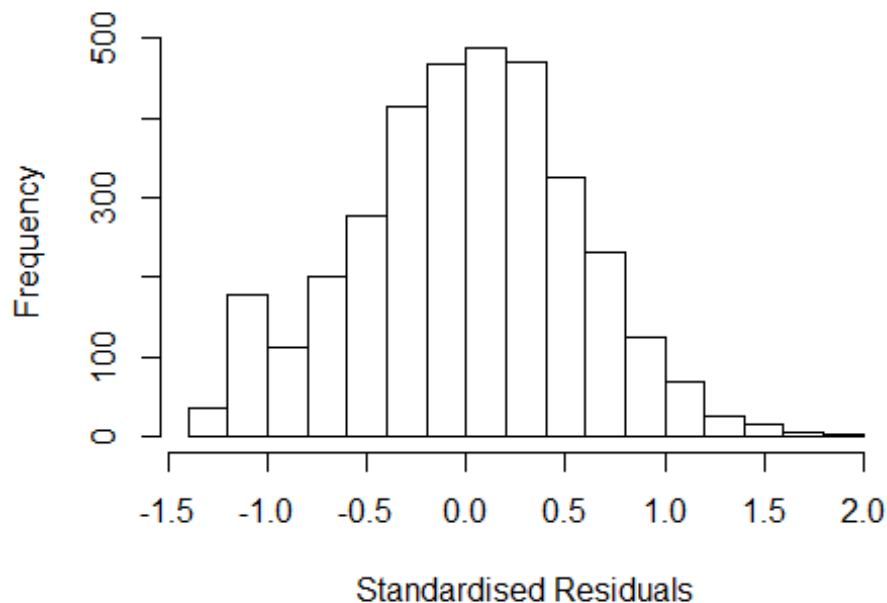
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2709 -0.3613 0.0122 0.3725 1.9241
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00226 0.04784 20.95 < 2e-16 ***
## FirstAuthorFemale1 0.02865 0.02612 1.10 0.27265
## LastAuthorFemale1 0.02560 0.02628 0.97 0.33009
## Year1997 -0.01262 0.06491 -0.19 0.84588
## Year1998 -0.03577 0.07059 -0.51 0.61241
## Year1999 0.04066 0.06727 0.60 0.54558
## Year2000 0.00255 0.06714 0.04 0.96972
## Year2001 -0.02659 0.06842 -0.39 0.69764
## Year2002 0.18842 0.06338 2.97 0.00297 **
## Year2003 0.20962 0.05956 3.52 0.00044 ***
## Year2004 0.21434 0.06087 3.52 0.00044 ***
## Year2005 0.10706 0.06279 1.71 0.08828 .
```

```

## Year2006          0.13605      0.06036      2.25  0.02425 *
## Year2007          0.13841      0.05773      2.40  0.01657 *
## Year2008          0.19843      0.06209      3.20  0.00141 **
## Year2009          0.13008      0.06183      2.10  0.03546 *
## Year2010          0.10017      0.06090      1.64  0.10007
## Year2011          0.13492      0.05998      2.25  0.02455 *
## Year2012          0.17365      0.06192      2.80  0.00507 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0204, Adjusted R-squared:  0.0153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 289 weights are ~= 1. The remaining 3146 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.206  0.863  0.952  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.014 1          1.007
## Year              1.014 16          1.000

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2635 -0.3601  0.0135  0.3711  1.9211
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.00389    0.04775   21.02  < 2e-16 ***
## FirstAuthorFemale1 0.04422    0.02142    2.06  0.03905 *
## Year1997      -0.01131    0.06484   -0.17  0.86154
## Year1998      -0.03525    0.07055   -0.50  0.61735
## Year1999       0.04142    0.06728    0.62  0.53813
## Year2000       0.00398    0.06714    0.06  0.95274
## Year2001      -0.02643    0.06857   -0.39  0.69993
## Year2002       0.19063    0.06327    3.01  0.00261 **
## Year2003       0.21162    0.05948    3.56  0.00038 ***
## Year2004       0.21541    0.06081    3.54  0.00040 ***
## Year2005       0.10801    0.06272    1.72  0.08512 .
## Year2006       0.13761    0.06026    2.28  0.02246 *
```

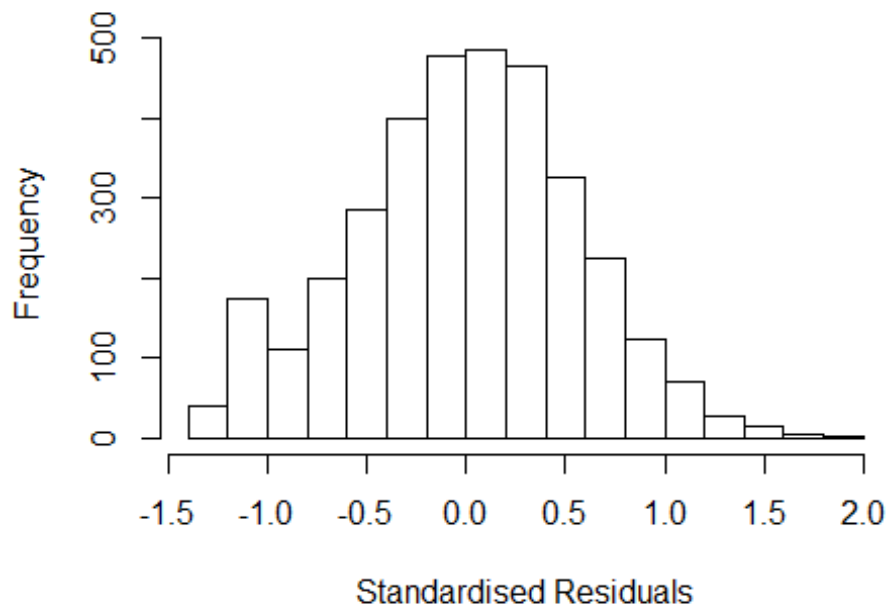


```

## Year2007          0.14031    0.05769    2.43  0.01506 *
## Year2008          0.19919    0.06211    3.21  0.00135 **
## Year2009          0.13212    0.06170    2.14  0.03232 *
## Year2010          0.10309    0.06068    1.70  0.08942 .
## Year2011          0.13663    0.05990    2.28  0.02262 *
## Year2012          0.17496    0.06189    2.83  0.00473 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0201, Adjusted R-squared:  0.0153
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 284 weights are ~= 1. The remaining 3151 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.208  0.863  0.952  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.023 1          1.011
## Year            1.023 16          1.001

```

## Residuals from last author



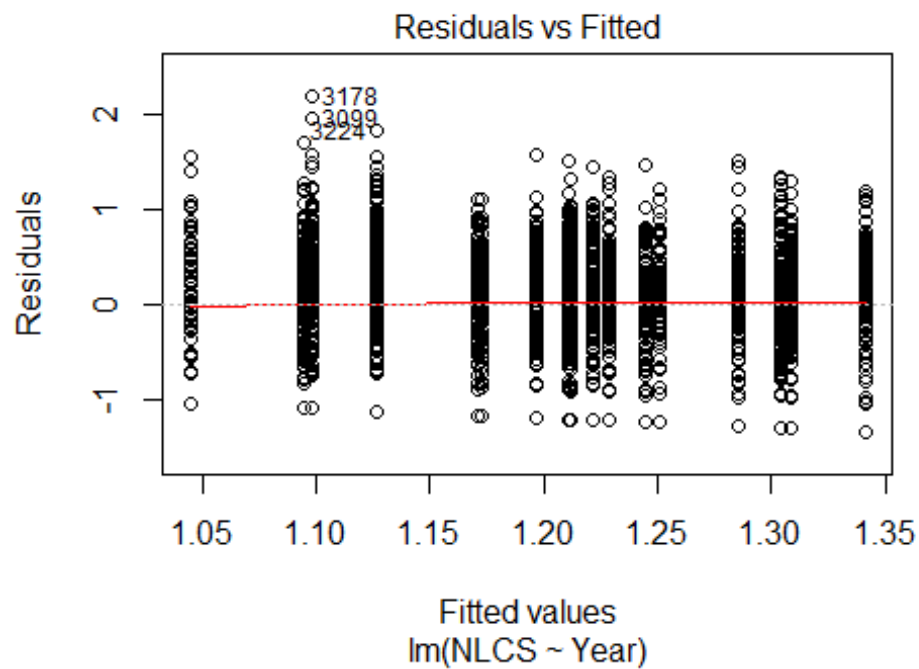
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26302 -0.35943 0.00999 0.37675 1.92064
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.00592 0.04778 21.05 < 2e-16 ***
## LastAuthorFemale1 0.04301 0.02154 2.00 0.04591 *
## Year1997 -0.01399 0.06496 -0.22 0.82944
## Year1998 -0.03725 0.07062 -0.53 0.59791
## Year1999 0.04027 0.06729 0.60 0.54960
## Year2000 0.00208 0.06717 0.03 0.97528
## Year2001 -0.02642 0.06844 -0.39 0.69949
## Year2002 0.18705 0.06335 2.95 0.00317 **
## Year2003 0.20840 0.05964 3.49 0.00048 ***
## Year2004 0.21409 0.06099 3.51 0.00045 ***
## Year2005 0.10650 0.06293 1.69 0.09063 .
## Year2006 0.13697 0.06046 2.27 0.02355 *
```

```

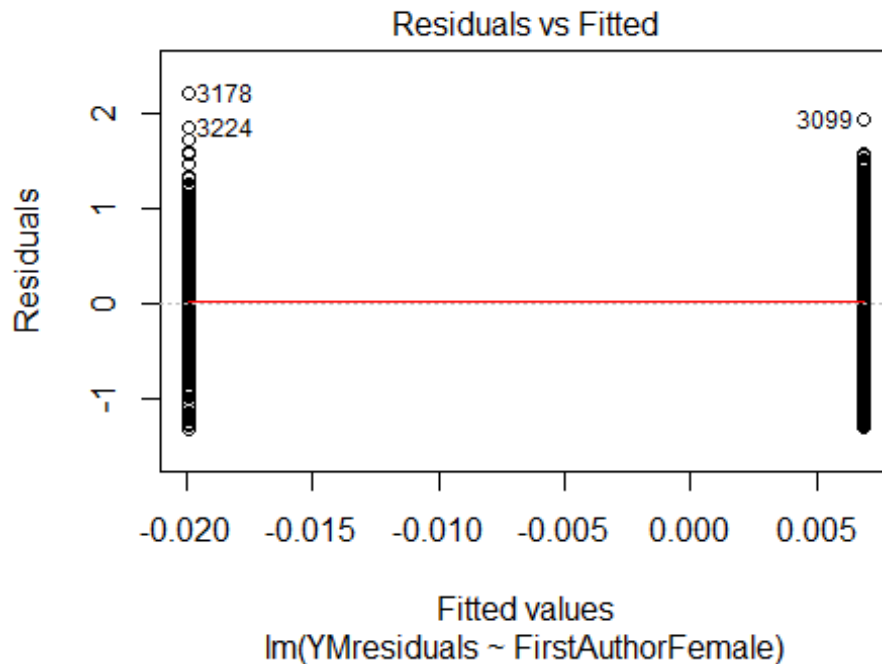
## Year2007      0.13807      0.05780      2.39  0.01697 *
## Year2008      0.19921      0.06214      3.21  0.00136 **
## Year2009      0.13052      0.06192      2.11  0.03511 *
## Year2010      0.09960      0.06100      1.63  0.10259
## Year2011      0.13501      0.06008      2.25  0.02469 *
## Year2012      0.17344      0.06203      2.80  0.00520 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.556
## Multiple R-squared:  0.0201, Adjusted R-squared:  0.0152
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 277 weights are ~= 1. The remaining 3158 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.207  0.863  0.953  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.91e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3435"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3322"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 134 115 156 185 227 203 192 222 176 148 145 150 163 200 224
## 2011 2012
## 267 307
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 95 115 137 108 71 165 195 155 130 131 129 139 170 192
## 2011 2012

```

```
## 237 267
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 108 93 111 133 105 65 160 190 149 124 127 123 135 163 183
## 2011 2012
## 231 256
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 48, df = 16, p-value = 5e-05
```

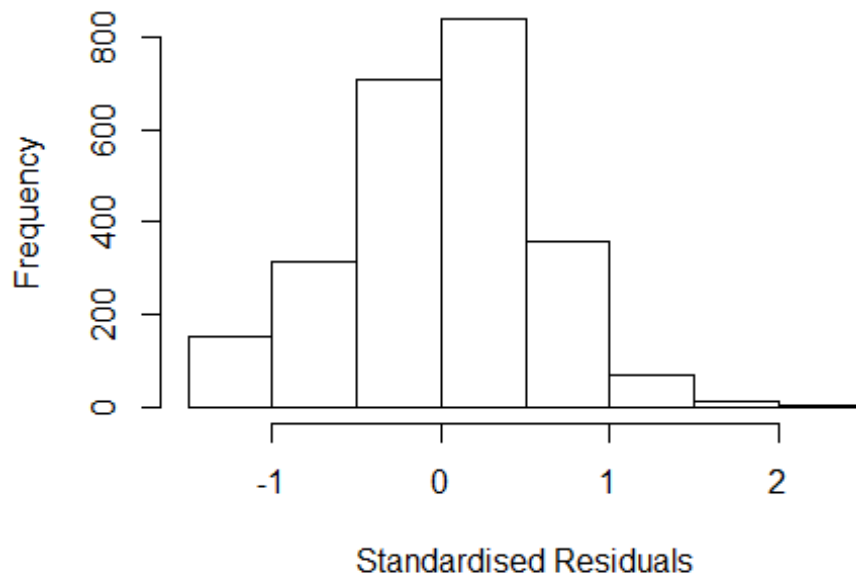


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.1, df = 1, p-value = 0.3
```



```
## [1] "Female first author team size 2018 geometric mean: 1.62073713540572"
## [1] "Male first author team size 2018 geometric mean: 1.45058159472733"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 5100, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.69308593869664"
## [1] "Male last author team size 2018 geometric mean: 1.40236676999684"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 6000, p-value = 0.007
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.560 1      1.249
## LastAuthorFemale  1.589 1      1.260
## UniqueAuthors    1.153 4      1.018
## Year             1.203 16      1.006
```

## Residuals from first and last author and team size



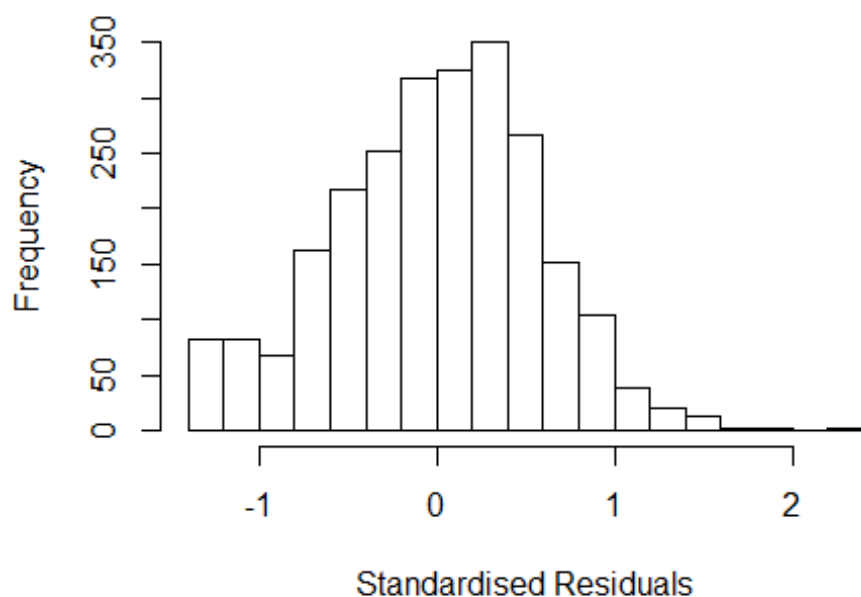
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4425 -0.3873 0.0272 0.3719 2.2958
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18898 0.06636 17.92 < 2e-16 ***
## FirstAuthorFemale1 -0.01970 0.03441 -0.57 0.567
## LastAuthorFemale1 -0.02713 0.03513 -0.77 0.440
## UniqueAuthors2 0.11198 0.02776 4.03 5.7e-05 ***
## UniqueAuthors3 0.24194 0.03801 6.36 2.3e-10 ***
## UniqueAuthors4 0.11270 0.06781 1.66 0.097 .
## UniqueAuthors5 0.29578 0.07454 3.97 7.5e-05 ***
## Year1997 0.04402 0.08712 0.51 0.613
## Year1998 -0.00407 0.08355 -0.05 0.961
## Year1999 -0.03645 0.07864 -0.46 0.643
```

```

## Year2000      0.01019    0.08773    0.12    0.908
## Year2001     -0.24114    0.10450   -2.31    0.021 *
## Year2002     -0.14852    0.08194   -1.81    0.070 .
## Year2003     -0.01064    0.07948   -0.13    0.894
## Year2004      0.08703    0.07616    1.14    0.253
## Year2005      0.01264    0.08213    0.15    0.878
## Year2006      0.03359    0.07832    0.43    0.668
## Year2007      0.10220    0.08270    1.24    0.217
## Year2008      0.01160    0.08340    0.14    0.889
## Year2009      0.07857    0.07883    1.00    0.319
## Year2010     -0.03225    0.07701   -0.42    0.675
## Year2011     -0.15091    0.07923   -1.90    0.057 .
## Year2012     -0.11156    0.07920   -1.41    0.159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.562
## Multiple R-squared:  0.0424, Adjusted R-squared:  0.0338
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 219 weights are ~= 1. The remaining 2237 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.058  0.869  0.949  0.904  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.570 1      1.253
## LastAuthorFemale  1.587 1      1.260
## Year              1.061 16      1.002

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3282 -0.3959  0.0275  0.3851  2.2507
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.238461   0.063971   19.36  <2e-16 ***
## FirstAuthorFemale1 -0.025915   0.034965   -0.74   0.459
## LastAuthorFemale1 -0.024542   0.035691   -0.69   0.492
## Year1997         0.042629   0.086169    0.49   0.621
## Year1998        -0.005048   0.081902   -0.06   0.951
## Year1999        -0.046723   0.077208   -0.61   0.545
## Year2000         0.022867   0.087193    0.26   0.793
## Year2001        -0.232260   0.104745   -2.22   0.027 *
## Year2002        -0.135983   0.081846   -1.66   0.097 .
## Year2003         0.000966   0.078274    0.01   0.990
## Year2004         0.087319   0.074186    1.18   0.239
## Year2005         0.031170   0.081241    0.38   0.701
```

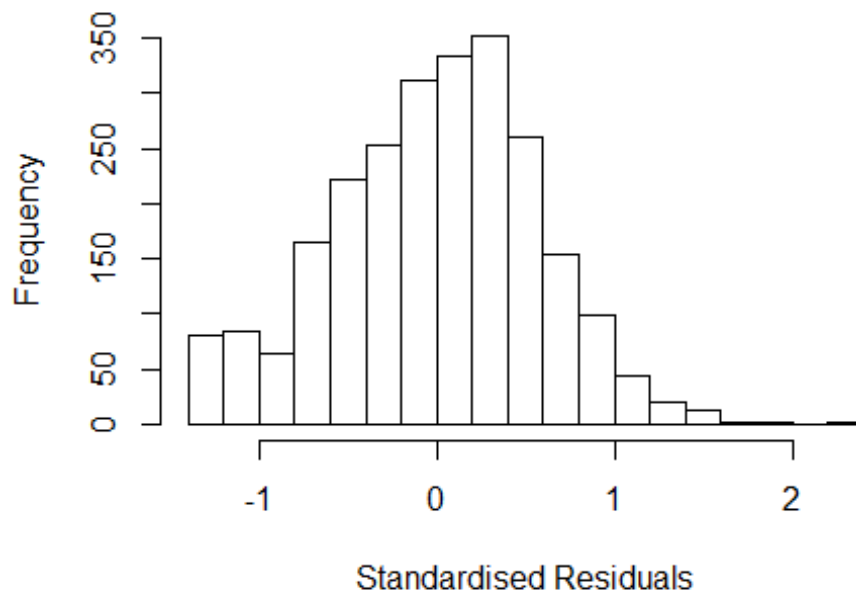


```

## Year2006          0.040804    0.077681    0.53    0.599
## Year2007          0.110307    0.081564    1.35    0.176
## Year2008          0.022959    0.082889    0.28    0.782
## Year2009          0.089765    0.076839    1.17    0.243
## Year2010         -0.023493    0.075327   -0.31    0.755
## Year2011         -0.151717    0.078492   -1.93    0.053 .
## Year2012         -0.106656    0.078152   -1.36    0.172
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.0232, Adjusted R-squared:  0.016
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 210 weights are ~= 1. The remaining 2246 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0867 0.8700 0.9490 0.9050 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.012
## Year              1.023 16      1.001

```

## Residuals from first author



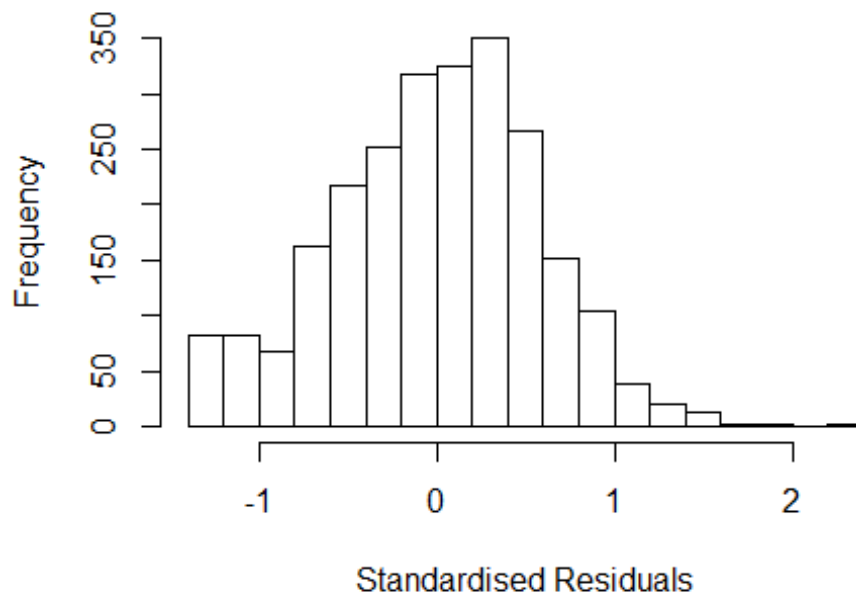
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3250 -0.3999 0.0291 0.3826 2.2444
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.236827 0.063954 19.34 <2e-16 ***
## FirstAuthorFemale1 -0.041945 0.028371 -1.48 0.139
## Year1997 0.043706 0.086236 0.51 0.612
## Year1998 -0.005225 0.081877 -0.06 0.949
## Year1999 -0.046807 0.077250 -0.61 0.545
## Year2000 0.022218 0.087206 0.25 0.799
## Year2001 -0.233157 0.104897 -2.22 0.026 *
## Year2002 -0.136656 0.081821 -1.67 0.095 .
## Year2003 0.000848 0.078207 0.01 0.991
## Year2004 0.087296 0.074187 1.18 0.239
## Year2005 0.032348 0.081214 0.40 0.690
## Year2006 0.041122 0.077669 0.53 0.597
```

```

## Year2007          0.111134    0.081533    1.36    0.173
## Year2008          0.020107    0.082582    0.24    0.808
## Year2009          0.088196    0.076775    1.15    0.251
## Year2010         -0.024551    0.075295   -0.33    0.744
## Year2011         -0.152327    0.078508   -1.94    0.052 .
## Year2012         -0.108897    0.078101   -1.39    0.163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.023, Adjusted R-squared:  0.0162
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 212 weights are ~= 1. The remaining 2244 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0887 0.8700 0.9490 0.9050 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.034 1          1.017
## Year            1.034 16          1.001

```

## Residuals from last author



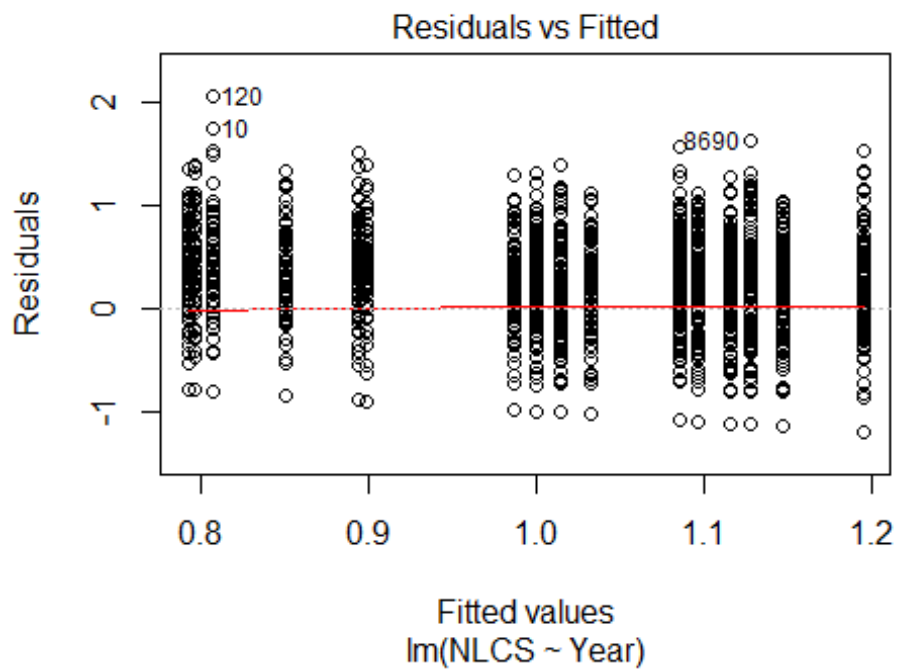
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3269 -0.3978 0.0299 0.3802 2.2443
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.23577 0.06387 19.35 <2e-16 ***
## LastAuthorFemale1 -0.04144 0.02889 -1.43 0.152
## Year1997 0.04311 0.08614 0.50 0.617
## Year1998 -0.00533 0.08191 -0.07 0.948
## Year1999 -0.04545 0.07718 -0.59 0.556
## Year2000 0.02474 0.08705 0.28 0.776
## Year2001 -0.22975 0.10464 -2.20 0.028 *
## Year2002 -0.13559 0.08200 -1.65 0.098 .
## Year2003 0.00113 0.07838 0.01 0.988
## Year2004 0.08721 0.07423 1.17 0.240
## Year2005 0.03022 0.08116 0.37 0.710
## Year2006 0.03967 0.07767 0.51 0.610
```

```

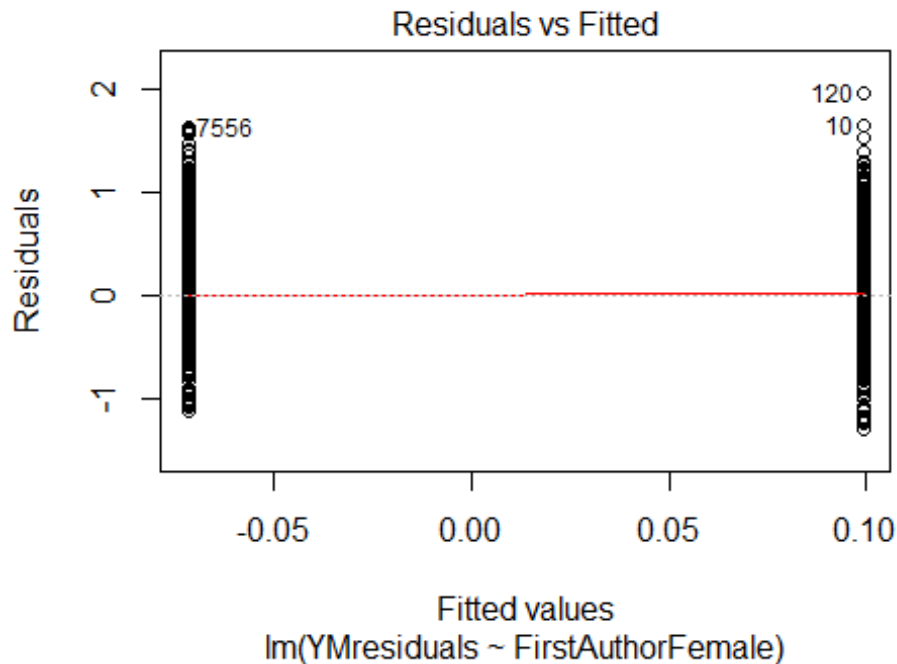
## Year2007      0.10965    0.08160    1.34    0.179
## Year2008      0.02447    0.08293    0.30    0.768
## Year2009      0.09108    0.07676    1.19    0.235
## Year2010     -0.02222    0.07529   -0.30    0.768
## Year2011     -0.15168    0.07845   -1.93    0.053 .
## Year2012     -0.10553    0.07820   -1.35    0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.572
## Multiple R-squared:  0.023, Adjusted R-squared:  0.0162
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 213 weights are ~= 1. The remaining 2243 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0886 0.8700 0.9490 0.9050 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.07e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2456"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3400"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 475 374 520 503 484 485 380 307 348 351 432 437 411 395 374
## 2011 2012
## 378 438
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 123 114 170 156 155 166 115 88 117 129 206 192 179 200 171
## 2011 2012

```

```
## 206 228
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 113 110 159 141 143 145 98 71 105 106 185 168 157 181 158
## 2011 2012
## 182 213
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 37, df = 16, p-value = 0.002
```

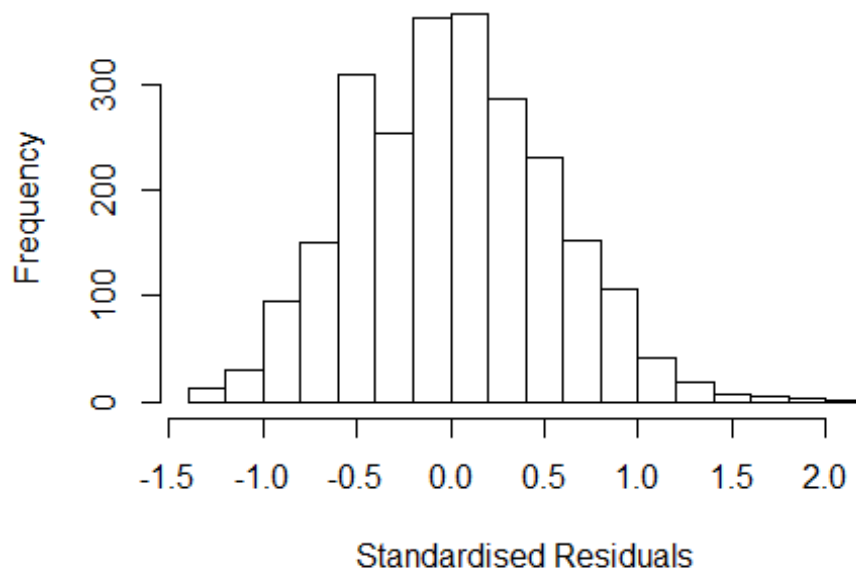


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 3, df = 1, p-value = 0.08
```



```
## [1] "Female first author team size 2018 geometric mean: 3.41705787803195"
## [1] "Male first author team size 2018 geometric mean: 3.10118963796682"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9300, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.26010489422943"
## [1] "Male last author team size 2018 geometric mean: 3.31459428755569"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 8900, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.221 1      1.105
## LastAuthorFemale  1.159 1      1.076
## UniqueAuthors    1.338 4      1.037
## Year             1.313 16      1.009
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.34734 -0.39013 0.00219 0.38057 2.04649
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.51805 0.07311 7.09 1.8e-12 ***
## FirstAuthorFemale1 0.11718 0.02474 4.74 2.3e-06 ***
## LastAuthorFemale1 0.00896 0.02670 0.34 0.737
## UniqueAuthors2 0.31330 0.04095 7.65 2.9e-14 ***
## UniqueAuthors3 0.50980 0.03745 13.61 < 2e-16 ***
## UniqueAuthors4 0.63718 0.03719 17.13 < 2e-16 ***
## UniqueAuthors5 0.65676 0.03254 20.18 < 2e-16 ***
## Year1997 0.02393 0.10020 0.24 0.811
## Year1998 0.00322 0.08557 0.04 0.970
## Year1999 0.01531 0.08614 0.18 0.859
```

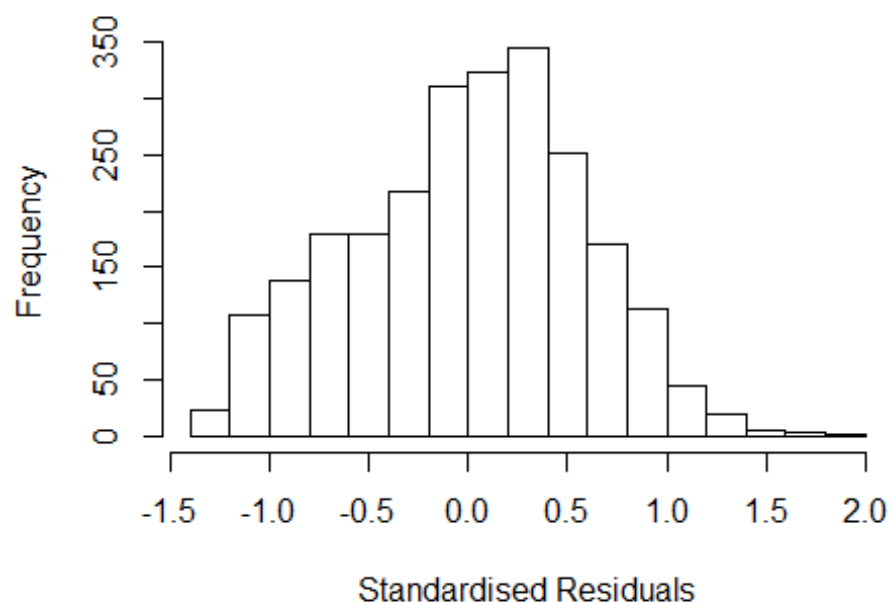


```

## Year2000      0.06338    0.08332    0.76    0.447
## Year2001     -0.02597    0.08522   -0.30    0.761
## Year2002      0.07493    0.09424    0.80    0.427
## Year2003      0.12246    0.09150    1.34    0.181
## Year2004      0.10209    0.09172    1.11    0.266
## Year2005      0.11382    0.08776    1.30    0.195
## Year2006      0.03058    0.08201    0.37    0.709
## Year2007     -0.00748    0.08240   -0.09    0.928
## Year2008      0.16431    0.08212    2.00    0.046 *
## Year2009      0.12257    0.07970    1.54    0.124
## Year2010      0.01617    0.08454    0.19    0.848
## Year2011      0.07003    0.08277    0.85    0.398
## Year2012      0.06432    0.08251    0.78    0.436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.25,   Adjusted R-squared:  0.243
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 194 weights are ~= 1. The remaining 2241 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0962 0.8720 0.9410 0.9060 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.11e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.134 1 1.065
## LastAuthorFemale 1.092 1 1.045
## Year 1.072 16 1.002

```

## Residuals from first and last author



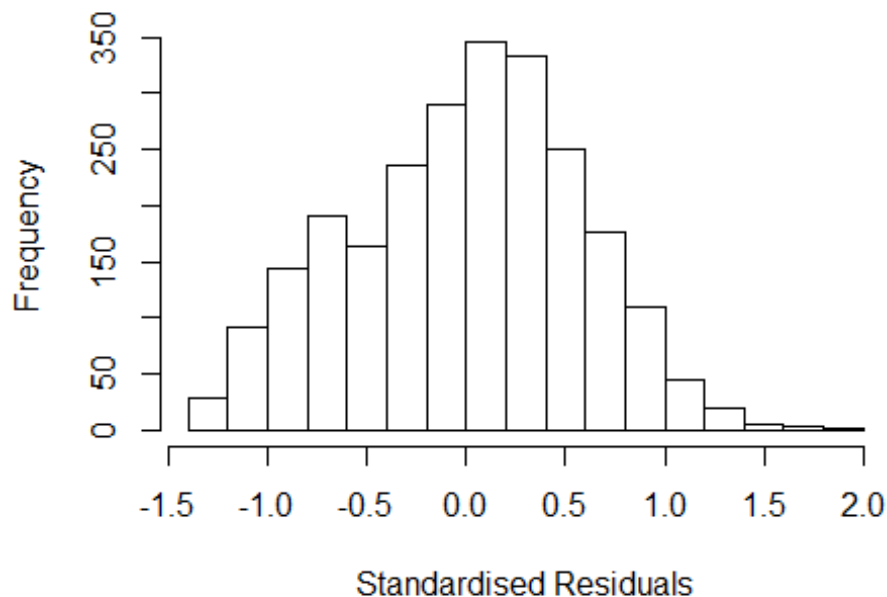
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.295 -0.419 0.043 0.400 1.919
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7122 0.0780 9.13 < 2e-16 ***
## FirstAuthorFemale1 0.2301 0.0262 8.77 < 2e-16 ***
## LastAuthorFemale1 -0.0406 0.0290 -1.40 0.16099
## Year1997 -0.0242 0.1052 -0.23 0.81831
## Year1998 0.0075 0.0940 0.08 0.93641
## Year1999 0.0423 0.0924 0.46 0.64665
## Year2000 0.1261 0.0899 1.40 0.16092
## Year2001 0.0786 0.0916 0.86 0.39102
## Year2002 0.1806 0.1011 1.79 0.07417 .
## Year2003 0.3277 0.1020 3.21 0.00133 **
## Year2004 0.2515 0.0980 2.57 0.01035 *
## Year2005 0.2948 0.0972 3.03 0.00243 **
```

```

## Year2006          0.1954      0.0913      2.14  0.03238 *
## Year2007          0.1827      0.0923      1.98  0.04792 *
## Year2008          0.3525      0.0909      3.88  0.00011 ***
## Year2009          0.3534      0.0867      4.08  4.7e-05 ***
## Year2010          0.2690      0.0942      2.85  0.00435 **
## Year2011          0.2927      0.0916      3.20  0.00141 **
## Year2012          0.3109      0.0916      3.39  0.00070 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.595
## Multiple R-squared:  0.0789, Adjusted R-squared:  0.072
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2250 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.276  0.866  0.948  0.912  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          4.11e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.058 1          1.029
## Year              1.058 16          1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2779 -0.4299 0.0497 0.4032 1.9375
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.70701 0.07802 9.06 < 2e-16 ***
## FirstAuthorFemale1 0.21652 0.02556 8.47 < 2e-16 ***
## Year1997 -0.02654 0.10522 -0.25 0.80088
## Year1998 0.00664 0.09400 0.07 0.94365
## Year1999 0.04102 0.09244 0.44 0.65726
## Year2000 0.12702 0.09006 1.41 0.15855
## Year2001 0.07871 0.09174 0.86 0.39099
## Year2002 0.18272 0.10137 1.80 0.07158 .
## Year2003 0.32568 0.10178 3.20 0.00139 **
## Year2004 0.25172 0.09788 2.57 0.01018 *
## Year2005 0.29319 0.09723 3.02 0.00259 **
## Year2006 0.19125 0.09131 2.09 0.03632 *
```

```

## Year2007          0.18144    0.09243    1.96  0.04975 *
## Year2008          0.35137    0.09096    3.86  0.00011 ***
## Year2009          0.35439    0.08677    4.08  4.6e-05 ***
## Year2010          0.26797    0.09431    2.84  0.00453 **
## Year2011          0.28966    0.09143    3.17  0.00155 **
## Year2012          0.30954    0.09168    3.38  0.00075 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.595
## Multiple R-squared:  0.0783, Adjusted R-squared:  0.0718
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 185 weights are ~= 1. The remaining 2250 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.267  0.866  0.947   0.912  0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.025 1      1.013
## Year              1.025 16      1.001
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.2092 -0.4204  0.0512  0.4149  2.1038

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.75722    0.07737   9.79 < 2e-16 ***
## LastAuthorFemale1 0.04859    0.02789   1.74  0.08162 .
## Year1997       -0.02204    0.10517  -0.21  0.83404
## Year1998        0.00683    0.09356   0.07  0.94177
## Year1999        0.05604    0.09199   0.61  0.54249
## Year2000        0.12985    0.09041   1.44  0.15108
## Year2001        0.09577    0.09192   1.04  0.29758
## Year2002        0.21632    0.10190   2.12  0.03387 *
## Year2003        0.35057    0.10209   3.43  0.00060 ***
## Year2004        0.27677    0.09710   2.85  0.00440 **
## Year2005        0.32410    0.09710   3.34  0.00086 ***
## Year2006        0.22513    0.09008   2.50  0.01252 *
## Year2007        0.22287    0.09158   2.43  0.01503 *
## Year2008        0.40341    0.09022   4.47  8.1e-06 ***
## Year2009        0.38559    0.08568   4.50  7.1e-06 ***
## Year2010        0.30165    0.09278   3.25  0.00117 **
## Year2011        0.32215    0.09075   3.55  0.00039 ***
## Year2012        0.35879    0.09141   3.93  8.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.605
## Multiple R-squared:  0.0512, Adjusted R-squared:  0.0445
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 183 weights are ~= 1. The remaining 2252 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.201  0.862  0.949  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.11e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 2435"
## [1] ""

```

```

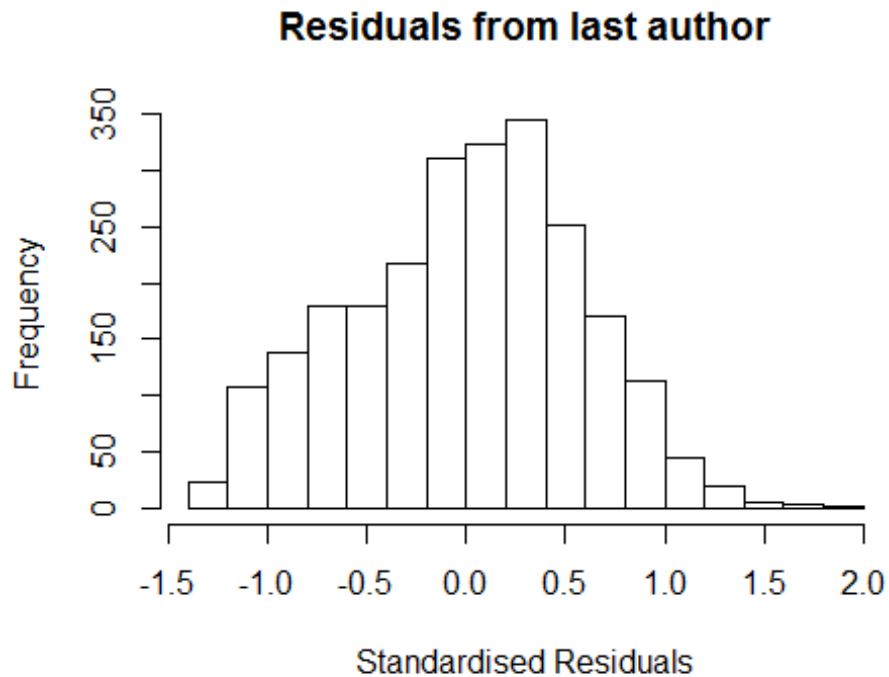
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3401"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 15 11 8 10 12 6 10 15 9 10 8 9 18 22 17
## 2011 2012
## 8 12
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 1 0 2 5 3 2 5 7 4 2 0 3 8 5
## 2011 2012
## 3 3
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 5 0 0 2 5 3 1 5 6 4 2 0 3 8 5
## 2011 2012
## 3 3
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.96193629594517"
## [1] "Male first author team size 2018 geometric mean: 2.28942848510666"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.34034731932072"
## [1] "Male last author team size 2018 geometric mean: 3.04683075789016"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

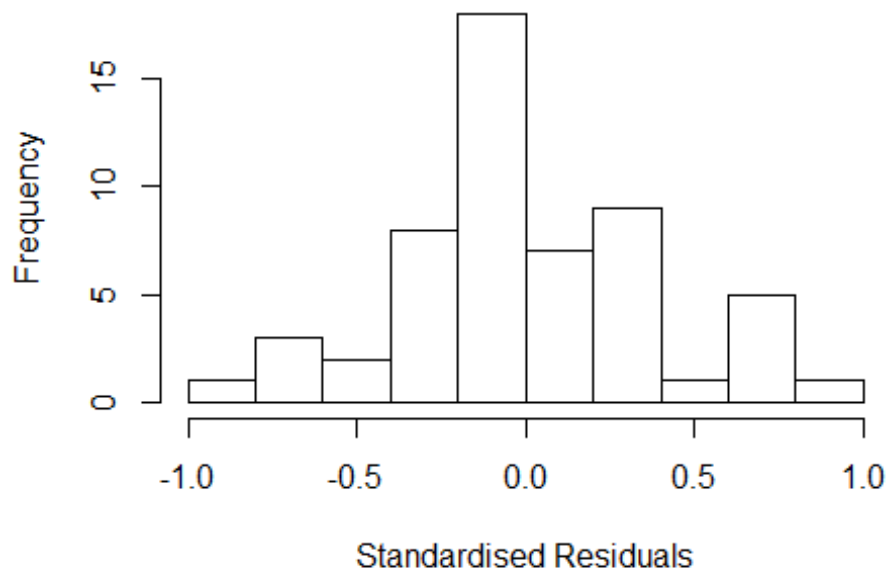
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 9, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##               GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 9.707e+01  1          9.852
## LastAuthorFemale  5.107e+00  1          2.260
## UniqueAuthors    2.433e+04  4          3.534
## Year              3.912e+05 13          1.641
```



## Residuals from first and last author and team size



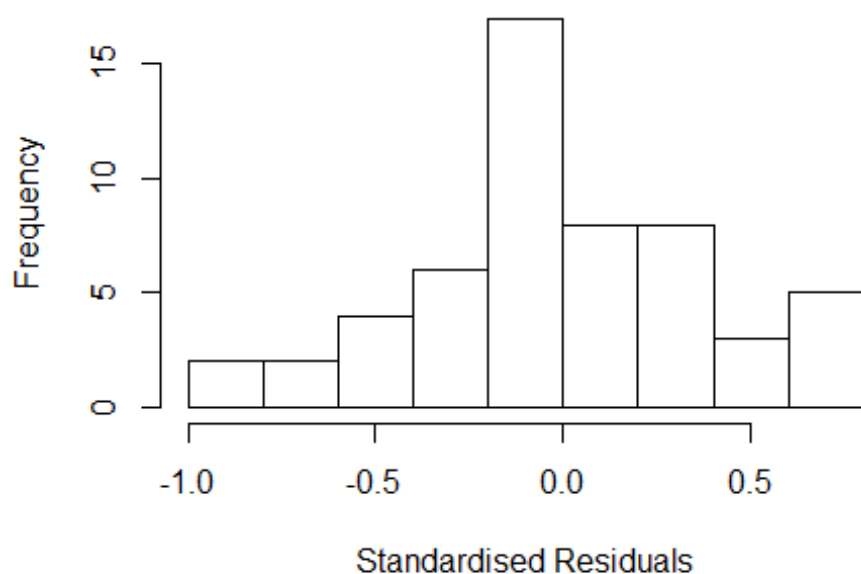
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9327 -0.2367 -0.0592 0.2556 0.8038
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.08265 0.44167 2.45 0.019 *
## FirstAuthorFemale1 -0.07103 0.24115 -0.29 0.770
## LastAuthorFemale1 0.16554 0.12274 1.35 0.186
## UniqueAuthors2 0.13886 0.20586 0.67 0.504
## UniqueAuthors3 0.15018 0.23086 0.65 0.520
## UniqueAuthors4 0.07858 0.24384 0.32 0.749
## UniqueAuthors5 0.37324 0.36932 1.01 0.319
## Year1999 -0.00665 0.44287 -0.02 0.988
## Year2000 -0.08247 0.35687 -0.23 0.819
## Year2001 -0.52294 0.40292 -1.30 0.203
```

```

## Year2002          0.64780    0.31601    2.05    0.048 *
## Year2003          -0.09612    0.41115   -0.23    0.817
## Year2004          -0.17786    0.35465   -0.50    0.619
## Year2005           0.19017    0.46427    0.41    0.685
## Year2006          -0.15412    0.33288   -0.46    0.646
## Year2008           0.06315    0.30709    0.21    0.838
## Year2009          -0.37529    0.30711   -1.22    0.230
## Year2010          -0.35088    0.32326   -1.09    0.285
## Year2011          -0.12575    0.94057   -0.13    0.894
## Year2012          -0.22664    0.35494   -0.64    0.527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.399
## Multiple R-squared:  0.267, Adjusted R-squared:  -0.131
## Convergence in 33 IRWLS iterations
##
## Robustness weights:
##  4 weights are ~= 1. The remaining 51 ones are summarized as
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.564  0.911  0.956   0.920   0.996   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.82e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 46.373 1      6.810
## LastAuthorFemale  6.359 1      2.522
## Year              218.486 13     1.230

```

## Residuals from first and last author



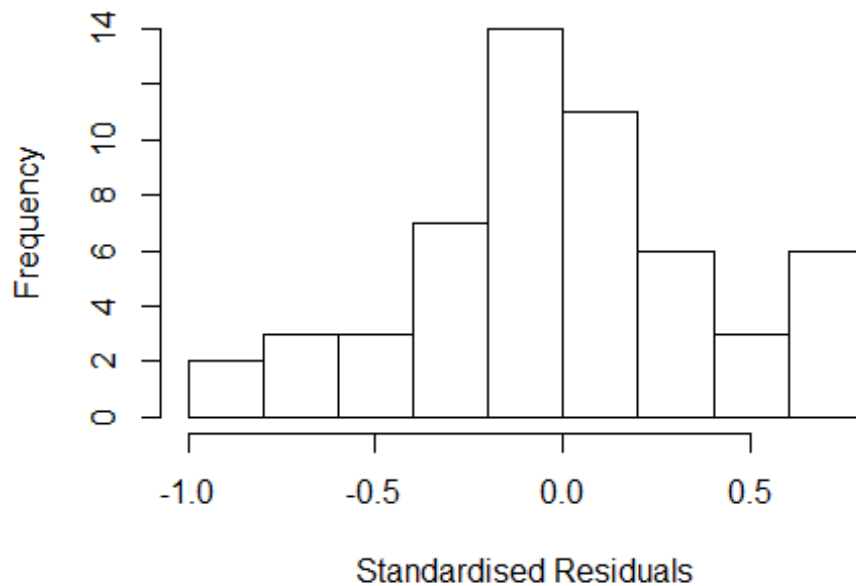
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.852 -0.193 -0.042 0.241 0.727
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.288 0.333 3.87 0.0004 ***
## FirstAuthorFemale1 -0.117 0.169 -0.69 0.4953
## LastAuthorFemale1 0.114 0.114 1.00 0.3238
## Year1999 -0.212 0.334 -0.63 0.5297
## Year2000 -0.219 0.321 -0.68 0.4998
## Year2001 -0.590 0.416 -1.42 0.1644
## Year2002 0.566 0.228 2.48 0.0176 *
## Year2003 -0.179 0.337 -0.53 0.5979
## Year2004 -0.274 0.337 -0.81 0.4225
## Year2005 0.104 0.332 0.31 0.7549
## Year2006 -0.203 0.265 -0.76 0.4489
## Year2008 0.108 0.349 0.31 0.7594
```

```

## Year2009          -0.320      0.308   -1.04   0.3056
## Year2010          -0.278      0.349   -0.80   0.4307
## Year2011          -0.148      0.441   -0.34   0.7392
## Year2012          -0.191      0.386   -0.49   0.6234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.384
## Multiple R-squared:  0.228, Adjusted R-squared:  -0.0692
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 3 weights are ~= 1. The remaining 52 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.603  0.853   0.968   0.911   0.990   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.82e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 48.89 1          6.992
## Year              48.89 13          1.161

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8964 -0.2041 -0.0284  0.2316  0.7878
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.3151     0.3368   3.90 0.00035 ***
## FirstAuthorFemale1 -0.0987     0.1715  -0.58 0.56831
## Year1999          -0.2391     0.3384  -0.71 0.48392
## Year2000          -0.2172     0.3404  -0.64 0.52703
## Year2001          -0.5757     0.3957  -1.45 0.15350
## Year2002           0.5216     0.2369   2.20 0.03353 *
## Year2003          -0.1582     0.3521  -0.45 0.65572
## Year2004          -0.2528     0.3437  -0.74 0.46638
## Year2005           0.1052     0.3395   0.31 0.75829
## Year2006          -0.1813     0.2889  -0.63 0.53389
## Year2008           0.1030     0.3785   0.27 0.78689
## Year2009          -0.3200     0.3180  -1.01 0.32034
```

```

## Year2010          -0.2508      0.3350   -0.75  0.45833
## Year2011          -0.1188      0.4328   -0.27  0.78510
## Year2012          -0.1734      0.3935   -0.44  0.66184
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.403
## Multiple R-squared:  0.207, Adjusted R-squared:  -0.0699
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 50 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.601  0.866  0.972  0.914  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.82e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 3.053 1          1.747
## Year              3.053 13          1.044
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -0.9112 -0.2167 -0.0192  0.2354  0.7208
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)

```

```

## (Intercept)          1.1782      0.2123      5.55      2e-06 ***
## LastAuthorFemale1    0.1031      0.1210      0.85      0.399
## Year1999             -0.1022      0.2148     -0.48      0.637
## Year2000             -0.1660      0.2574     -0.65      0.523
## Year2001             -0.5902      0.3975     -1.48      0.145
## Year2002              0.5598      0.2123      2.64      0.012 *
## Year2003             -0.1370      0.3195     -0.43      0.671
## Year2004             -0.2100      0.2590     -0.81      0.422
## Year2005              0.2170      0.2164      1.00      0.322
## Year2006             -0.1452      0.2193     -0.66      0.511
## Year2008              0.1304      0.3067      0.43      0.673
## Year2009             -0.2670      0.2535     -1.05      0.299
## Year2010             -0.2265      0.3370     -0.67      0.505
## Year2011             -0.0458      0.4056     -0.11      0.911
## Year2012             -0.1163      0.3586     -0.32      0.747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.386
## Multiple R-squared:  0.213, Adjusted R-squared:  -0.0626
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 5 weights are ~= 1. The remaining 50 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.557  0.835   0.966   0.907   0.989   0.999
## Algorithmic parameters:
##           tuning.chi              bb           tuning.psi           refine.tol
##           1.55e+00              5.00e-01           4.69e+00           1.00e-07
##           rel.tol              solve.tol           eps.outlier           eps.x
##           1.00e-07              1.00e-07           1.82e-03           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01              5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##           0          1000          0
##           psi              subsampling              cov
##           "bisquare"              "nonsingular"              ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis:  55"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3402"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"

```

```

##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   51   60   41   55   40   43   56   57   42   34   46   35   37   78   70
## 2011 2012
##   47   65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   11   13    9    7    2    4    0    2    3    6    2    2    2    3    3
## 2011 2012
##    1    2
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    5    8    4    3    1    3    0    1    3    5    2    2    1    3    2
## 2011 2012
##    1    2
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.41565251562535"
## [1] "Male first author team size 2018 geometric mean: 6.64939976115097"

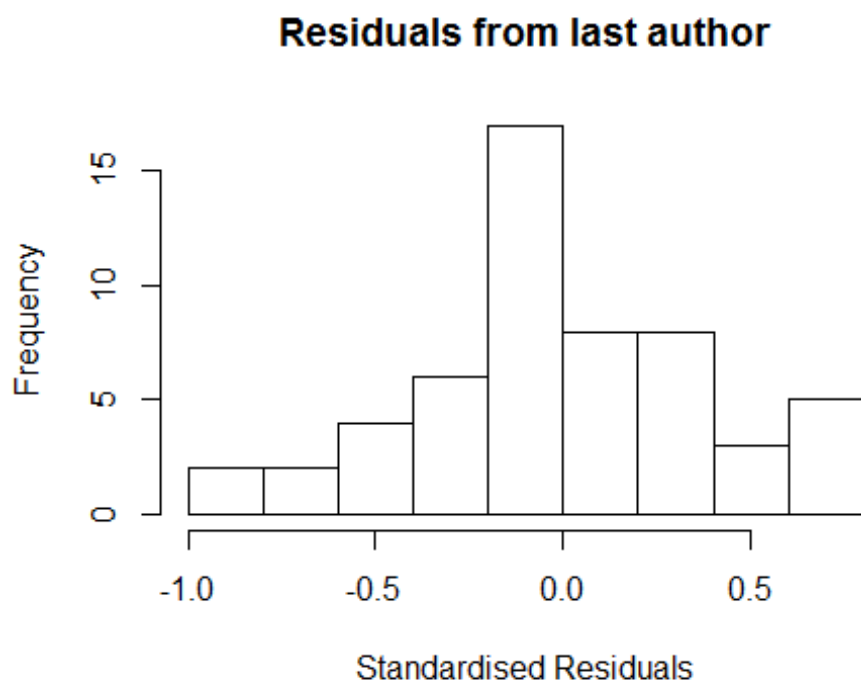
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.45723364049189"
## [1] "Male last author team size 2018 geometric mean: 2"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

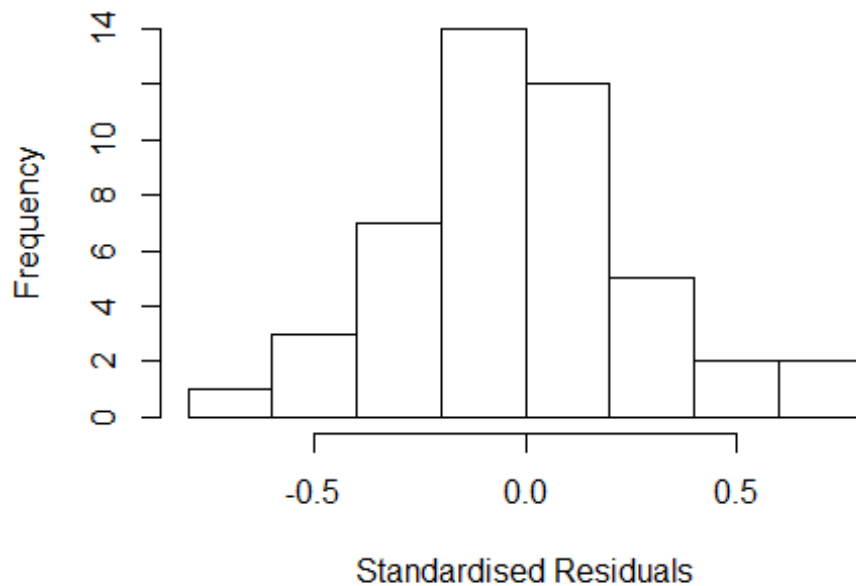
```





```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 7, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.282e+02 1      15.105
## LastAuthorFemale -2.975e-10 1         NaN
## UniqueAuthors    -1.847e+04 3         NaN
## Year              3.944e+18 15         4.167
```

## Residuals from first and last author and team size



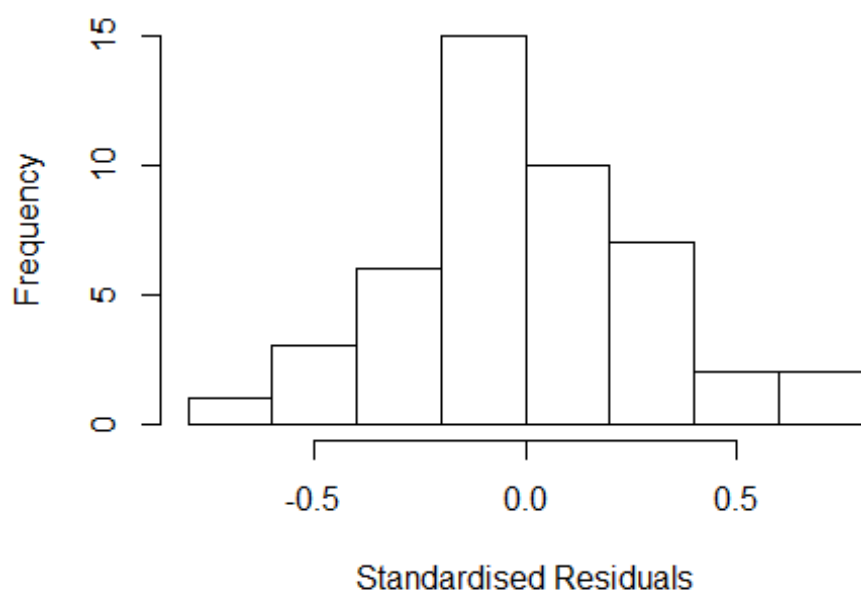
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -7.57e-01 -1.60e-01 2.22e-16 1.33e-01 7.55e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1885 0.2375 5.00 3.7e-05 ***
## FirstAuthorFemale1 -0.1495 0.1187 -1.26 0.21945
## LastAuthorFemale1 -0.0178 0.1403 -0.13 0.89994
## UniqueAuthors2 -0.0122 0.2452 -0.05 0.96064
## UniqueAuthors3 0.2839 0.1029 2.76 0.01072 *
## UniqueAuthors5 0.2779 0.1188 2.34 0.02758 *
## Year1997 -0.4195 0.2475 -1.69 0.10254
## Year1998 -0.2202 0.3364 -0.65 0.51876
## Year1999 -0.1432 0.2627 -0.55 0.59046
## Year2000 -0.3965 0.2375 -1.67 0.10745
```

```

## Year2001          -0.7608      0.2691    -2.83   0.00910 **
## Year2003          -0.1739      0.2895    -0.60   0.55353
## Year2004          -0.4317      0.6496    -0.66   0.51240
## Year2005           0.2570      0.2618     0.98   0.33574
## Year2006          -0.1739      0.2450    -0.71   0.48440
## Year2007           0.4584      0.2269     2.02   0.05416 .
## Year2008          -0.5355      0.2375    -2.26   0.03314 *
## Year2009          -0.3072      0.4201    -0.73   0.47140
## Year2010          -0.9906      0.2498    -3.97   0.00054 ***
## Year2011          -1.1885      0.2375    -5.00   3.7e-05 ***
## Year2012          -0.9402      0.2413    -3.90   0.00065 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.388
## Multiple R-squared:  0.605, Adjusted R-squared:  0.288
## Convergence in 23 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 31 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.683  0.911  0.962  0.924  0.985  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.17e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.888e+14 1      1.374e+07
## LastAuthorFemale  1.750e+02 1      1.323e+01
## Year              3.250e+16 15      3.551e+00

```

## Residuals from first and last author



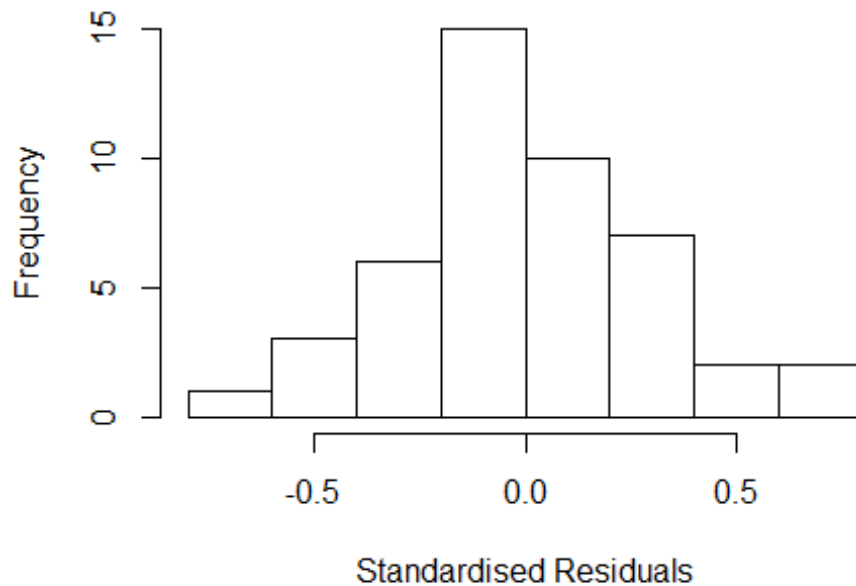
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -7.55e-01 -1.88e-01 -2.22e-16 1.82e-01 7.45e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1431 0.2412 4.74 5.7e-05 ***
## FirstAuthorFemale1 -0.1505 0.1207 -1.25 0.2225
## LastAuthorFemale1 0.0387 0.1337 0.29 0.7745
## Year1997 -0.3777 0.2416 -1.56 0.1292
## Year1998 -0.2340 0.3297 -0.71 0.4837
## Year1999 -0.1410 0.2598 -0.54 0.5917
## Year2000 -0.3511 0.2412 -1.46 0.1567
## Year2001 -0.5399 0.2825 -1.91 0.0663 .
## Year2003 0.1565 0.2558 0.61 0.5456
## Year2004 -0.3884 0.5734 -0.68 0.5038
## Year2005 0.5016 0.2442 2.05 0.0494 *
## Year2006 -0.0455 0.2340 -0.19 0.8473
```

```

## Year2007          0.4423      0.2279      1.94      0.0624 .
## Year2008          -0.4901      0.2412     -2.03      0.0518 .
## Year2009          -0.2592      0.3924     -0.66      0.5142
## Year2010          -0.8088      0.2762     -2.93      0.0067 **
## Year2011          -1.1431      0.2412     -4.74      5.7e-05 ***
## Year2012          -0.6386      0.2295     -2.78      0.0095 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.425
## Multiple R-squared:  0.568, Adjusted R-squared:  0.306
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 35 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.733  0.928  0.969  0.940  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.17e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 266 1          16.309
## Year              266 15          1.205

```

## Residuals from first author



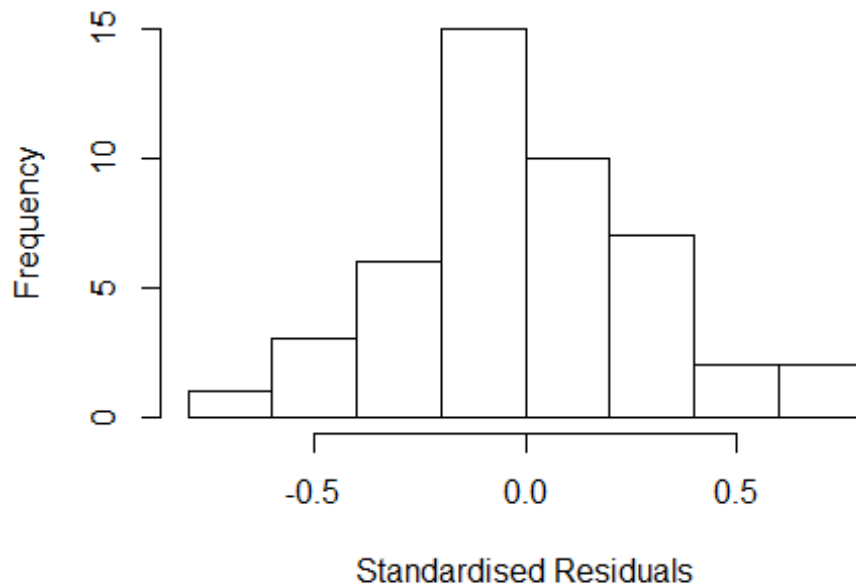
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.55e-01 -1.90e-01  3.47e-17  1.95e-01  7.45e-01
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1531     0.2368   4.87 3.6e-05 ***
## FirstAuthorFemale1 -0.1249     0.0999  -1.25  0.2211
## Year1997         -0.3781     0.2428  -1.56  0.1303
## Year1998         -0.2310     0.3326  -0.69  0.4930
## Year1999         -0.1492     0.2726  -0.55  0.5883
## Year2000         -0.3611     0.2368  -1.52  0.1381
## Year2001         -0.5547     0.2723  -2.04  0.0509 .
## Year2003          0.1208     0.2288   0.53  0.6014
## Year2004         -0.3981     0.5927  -0.67  0.5071
## Year2005          0.4868     0.2410   2.02  0.0527 .
## Year2006         -0.0296     0.2303  -0.13  0.8986
## Year2007          0.4453     0.2289   1.95  0.0614 .
```

```

## Year2008          -0.5001      0.2368   -2.11    0.0434 *
## Year2009          -0.2723      0.3933   -0.69    0.4941
## Year2010          -0.8316      0.2698   -3.08    0.0045 **
## Year2011          -1.1531      0.2368   -4.87    3.6e-05 ***
## Year2012          -0.6421      0.2291   -2.80    0.0089 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.571, Adjusted R-squared:  0.334
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 34 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.715  0.924  0.966  0.934  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.17e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 226.4  1          15.046
## Year            226.4 15          1.198

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -7.55e-01 -1.61e-01 -1.16e-15 2.27e-01 7.45e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1021 0.2565 4.30 0.00018 ***
## LastAuthorFemale1 -0.0607 0.1130 -0.54 0.59523
## Year1997 -0.3744 0.2527 -1.48 0.14933
## Year1998 -0.2442 0.3390 -0.72 0.47698
## Year1999 -0.1804 0.2899 -0.62 0.53876
## Year2000 -0.3101 0.2565 -1.21 0.23634
## Year2001 -0.5700 0.2820 -2.02 0.05251 .
## Year2003 0.0469 0.2565 0.18 0.85630
## Year2004 -0.3472 0.5991 -0.58 0.56670
## Year2005 0.4631 0.2525 1.83 0.07690 .
## Year2006 0.0196 0.2518 0.08 0.93862
## Year2007 0.4321 0.2382 1.81 0.08012 .
```

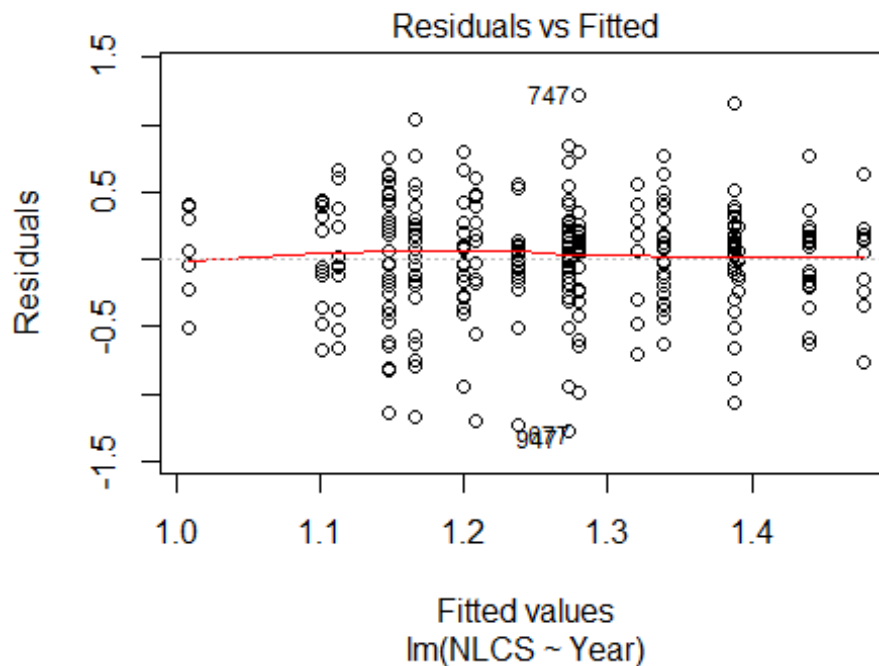


```

## Year2008          -0.4491      0.2565   -1.75  0.09049 .
## Year2009          -0.2211      0.4052   -0.55  0.58941
## Year2010          -0.8431      0.3238   -2.60  0.01439 *
## Year2011          -1.1021      0.2565   -4.30  0.00018 ***
## Year2012          -0.6233      0.2462   -2.53  0.01703 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.565, Adjusted R-squared:  0.325
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 35 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.716  0.914  0.967  0.935  0.987  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.17e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 46"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3403"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   73   63   71   54   51   51   45   53   36   54   42   38   46   70   75
## 2011 2012
##   67   65
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12   15   10   11    8   11    8   17   11   19   20   18   24   28   30
## 2011 2012
##   38   31

```

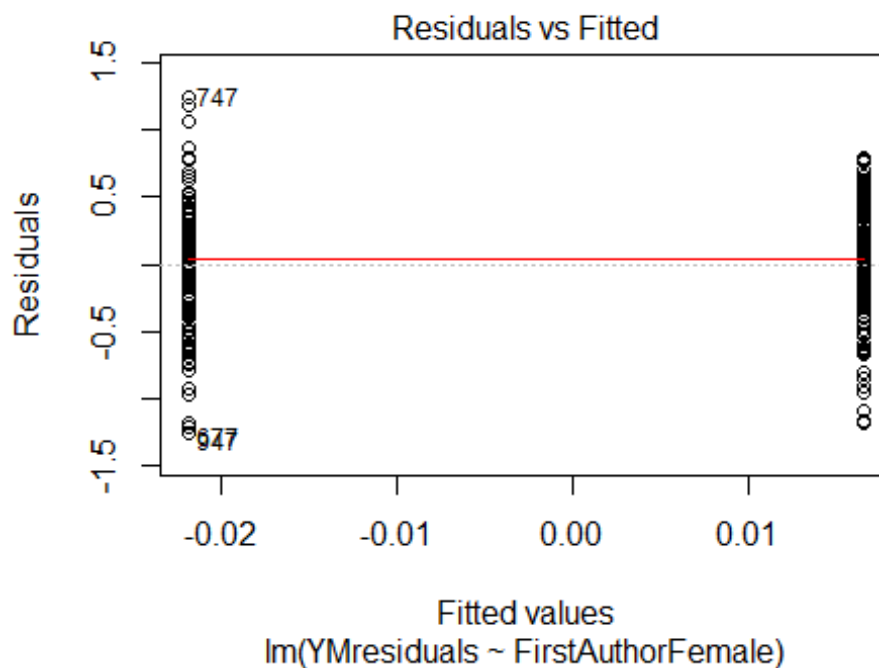
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 11 13 9 8 5 8 7 15 10 15 18 16 24 27 26
## 2011 2012
## 36 28
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 25, df = 16, p-value = 0.06
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.6, df = 1, p-value = 0.1
## [1] "Female first author team size 2018 geometric mean: 4.13635717862688"
## [1] "Male first author team size 2018 geometric mean: 3.23621267495743"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
```

```
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 140, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.06672906964239"
## [1] "Male last author team size 2018 geometric mean: 3.6416326425641"

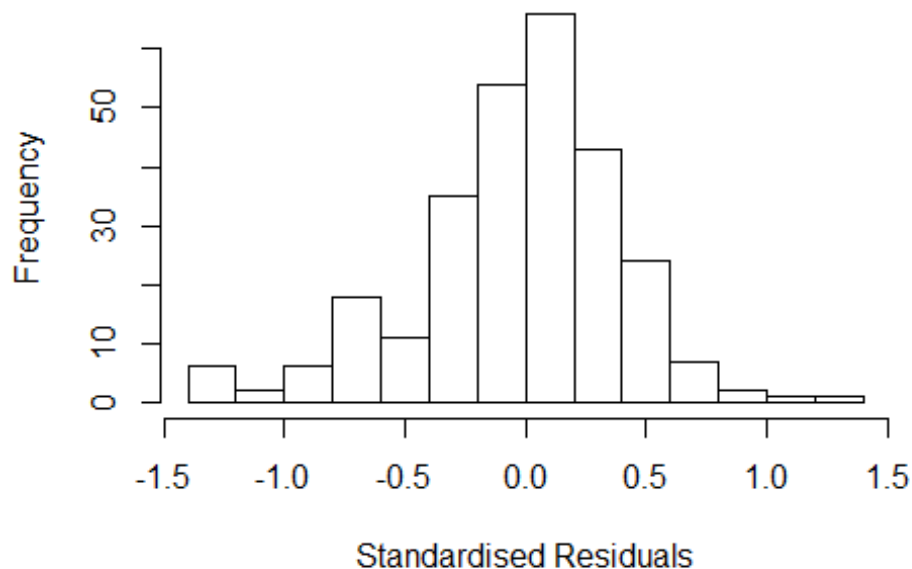
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 150, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVI	F	Df	GVI <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.815	1		1.347
LastAuthorFemale	2.302	1		1.517
UniqueAuthors	5.586	4		1.240
Year	17.916	16		1.094

## Residuals from first and last author and team size



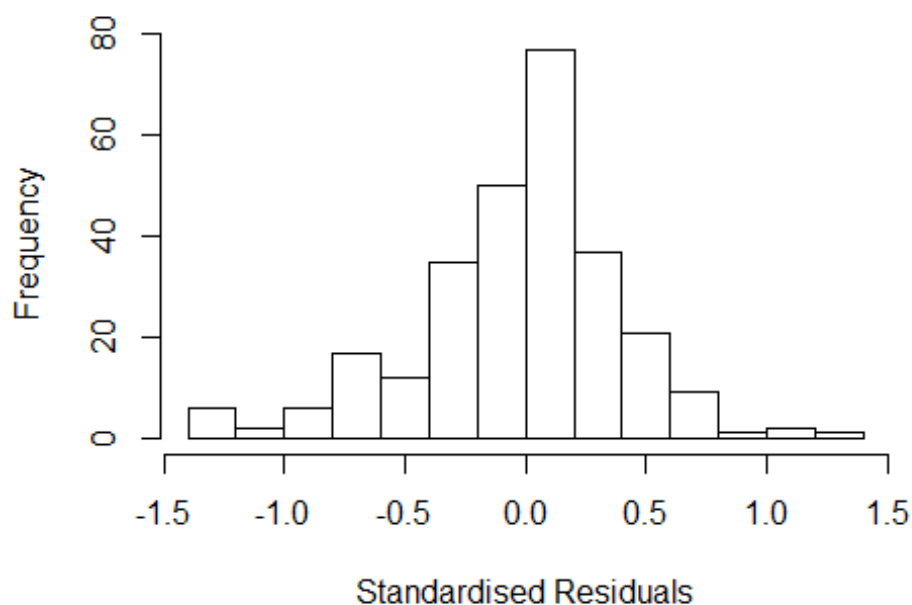
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2857 -0.2496 0.0245 0.2264 1.2541
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0925 0.1816 6.02 6.2e-09 ***
## FirstAuthorFemale1 0.0579 0.0577 1.00 0.3168
## LastAuthorFemale1 0.0441 0.0663 0.67 0.5058
## UniqueAuthors2 0.0506 0.1132 0.45 0.6552
## UniqueAuthors3 -0.0385 0.1098 -0.35 0.7260
## UniqueAuthors4 0.0400 0.1388 0.29 0.7736
## UniqueAuthors5 0.0408 0.1197 0.34 0.7335
## Year1997 0.1817 0.2115 0.86 0.3911
## Year1998 -0.1162 0.1979 -0.59 0.5578
## Year1999 0.3783 0.2016 1.88 0.0617 .
```

```

## Year2000          0.4787      0.1745      2.74      0.0065 **
## Year2001          0.0209      0.2341      0.09      0.9291
## Year2002          0.3298      0.1595      2.07      0.0397 *
## Year2003          0.0517      0.1780      0.29      0.7716
## Year2004          0.1104      0.1772      0.62      0.5340
## Year2005          0.2482      0.1862      1.33      0.1838
## Year2006          0.1680      0.1627      1.03      0.3028
## Year2007          0.1524      0.2048      0.74      0.4573
## Year2008          0.1571      0.1838      0.85      0.3935
## Year2009          0.2434      0.1813      1.34      0.1806
## Year2010          0.1592      0.1726      0.92      0.3571
## Year2011          0.0554      0.1954      0.28      0.7771
## Year2012          0.0504      0.1839      0.27      0.7841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.353
## Multiple R-squared:  0.0956, Adjusted R-squared:  0.017
## Convergence in 20 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 253 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.156  0.863  0.951  0.878  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      3.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.819 1      1.349
## LastAuthorFemale  2.228 1      1.493
## Year              3.852 16      1.043

```

## Residuals from first and last author



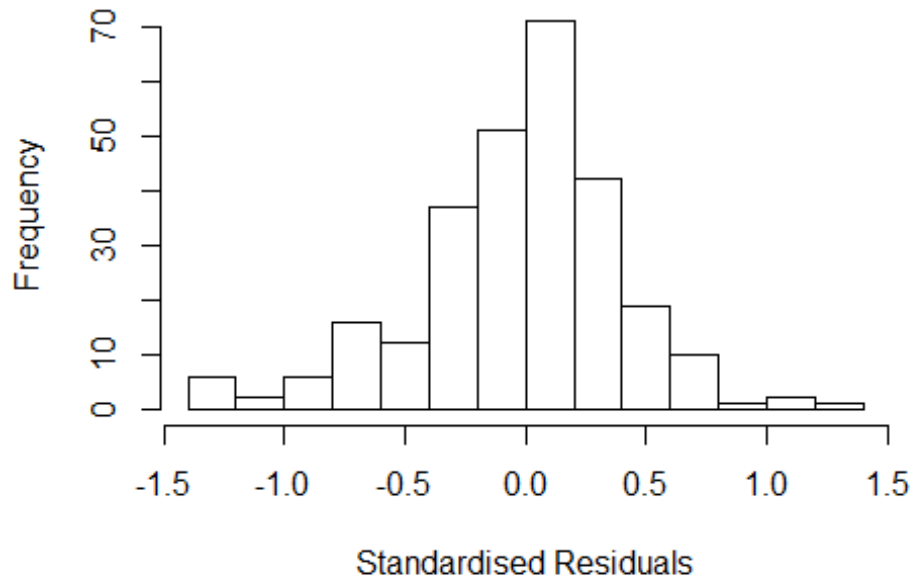
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2889 -0.2321  0.0317  0.2213  1.2326
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.0985     0.1394   7.88 9.2e-14 ***
## FirstAuthorFemale1  0.0638     0.0582   1.10  0.2738
## LastAuthorFemale1  0.0383     0.0662   0.58  0.5634
## Year1997          0.1896     0.2118   0.90  0.3715
## Year1998         -0.0971     0.1826  -0.53  0.5954
## Year1999          0.3746     0.2029   1.85  0.0661 .
## Year2000          0.4851     0.1604   3.02  0.0027 **
## Year2001          0.0120     0.2380   0.05  0.9598
## Year2002          0.3061     0.1519   2.02  0.0449 *
## Year2003          0.0622     0.1686   0.37  0.7124
## Year2004          0.1498     0.1630   0.92  0.3591
## Year2005          0.2475     0.1778   1.39  0.1651
```

```

## Year2006          0.1637      0.1525      1.07      0.2840
## Year2007          0.1679      0.1979      0.85      0.3972
## Year2008          0.1748      0.1739      1.01      0.3155
## Year2009          0.2540      0.1728      1.47      0.1429
## Year2010          0.1699      0.1634      1.04      0.2994
## Year2011          0.0882      0.1844      0.48      0.6328
## Year2012          0.0535      0.1771      0.30      0.7626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.347
## Multiple R-squared:  0.0907, Adjusted R-squared:  0.027
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 260 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.137  0.856  0.953  0.876  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.826 1      1.351
## Year              1.826 16      1.019

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2961 -0.2316 0.0356 0.2155 1.2276
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.1030 0.1412 7.81 1.4e-13 ***
## FirstAuthorFemale1 0.0684 0.0584 1.17 0.2432
## Year1997 0.1932 0.2061 0.94 0.3495
## Year1998 -0.0925 0.1790 -0.52 0.6058
## Year1999 0.3670 0.2044 1.80 0.0738 .
## Year2000 0.4868 0.1588 3.07 0.0024 **
## Year2001 0.0154 0.2344 0.07 0.9475
## Year2002 0.3005 0.1541 1.95 0.0524 .
## Year2003 0.0615 0.1684 0.36 0.7155
## Year2004 0.1495 0.1614 0.93 0.3553
## Year2005 0.2647 0.1660 1.59 0.1119
## Year2006 0.1657 0.1525 1.09 0.2781
```

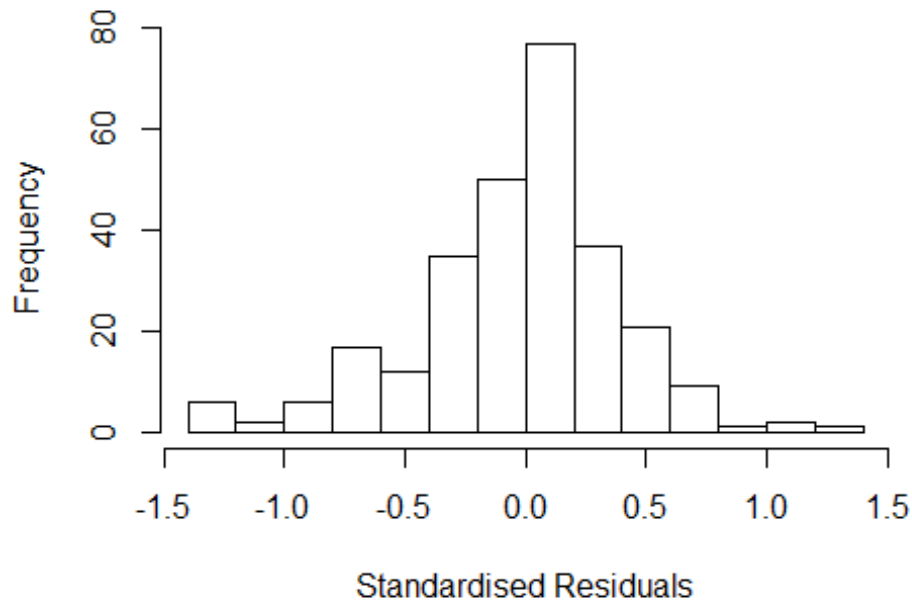


```

## Year2007          0.1684      0.1968      0.86      0.3929
## Year2008          0.1811      0.1715      1.06      0.2918
## Year2009          0.2678      0.1656      1.62      0.1072
## Year2010          0.1751      0.1605      1.09      0.2762
## Year2011          0.0992      0.1746      0.57      0.5703
## Year2012          0.0595      0.1749      0.34      0.7340
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.354
## Multiple R-squared:  0.0881, Adjusted R-squared:  0.028
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 262 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.151  0.859   0.958   0.881   0.988   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 2.322 1      1.524
## Year              2.322 16      1.027

```

## Residuals from last author



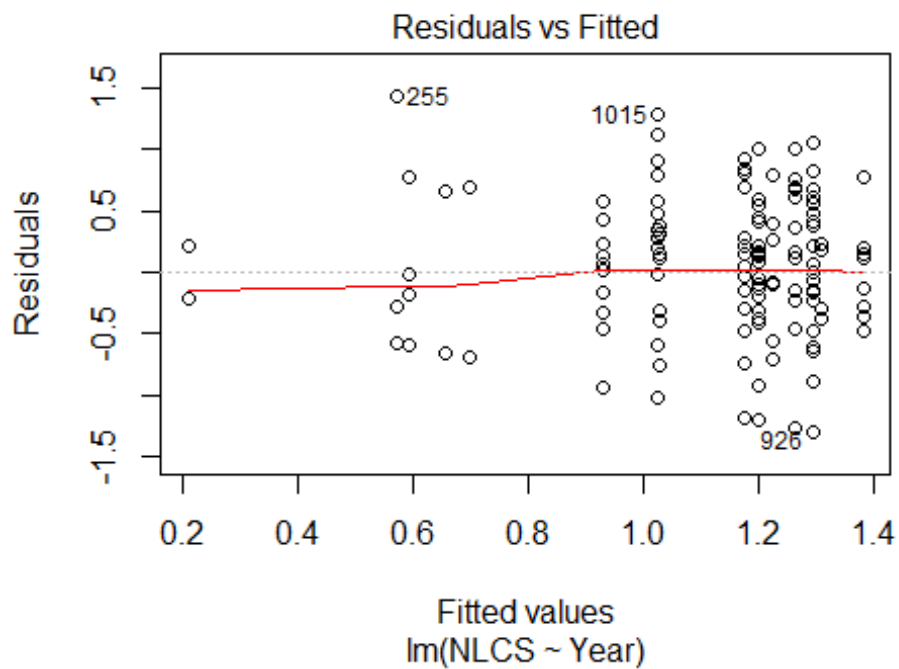
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.3177 -0.2418  0.0236  0.2043  1.1869
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.1039     0.1427   7.73 2.4e-13 ***
## LastAuthorFemale1  0.0460     0.0676   0.68  0.4968
## Year1997          0.2096     0.2115   0.99  0.3228
## Year1998         -0.0798     0.1836  -0.43  0.6643
## Year1999          0.4019     0.2025   1.98  0.0482 *
## Year2000          0.5036     0.1675   3.01  0.0029 **
## Year2001          0.0139     0.2459   0.06  0.9550
## Year2002          0.3193     0.1532   2.08  0.0381 *
## Year2003          0.0911     0.1738   0.52  0.6005
## Year2004          0.1696     0.1663   1.02  0.3087
## Year2005          0.2920     0.1760   1.66  0.0984 .
## Year2006          0.1801     0.1571   1.15  0.2528
```

```

## Year2007          0.2082      0.1940      1.07      0.2843
## Year2008          0.2165      0.1723      1.26      0.2101
## Year2009          0.2985      0.1681      1.78      0.0770 .
## Year2010          0.2138      0.1625      1.32      0.1894
## Year2011          0.1111      0.1881      0.59      0.5553
## Year2012          0.0919      0.1773      0.52      0.6048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.345
## Multiple R-squared:  0.0867, Adjusted R-squared:  0.0265
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 258 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.112  0.846  0.954  0.874  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      3.62e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 276"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3404"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   51   55   42   66   57   66   57   58   45   47   52   62   42   69   66
## 2011 2012
##   61   76
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    2    5    0    2    4    2    4    7    5   15    9   11    8   18   25
## 2011 2012

```

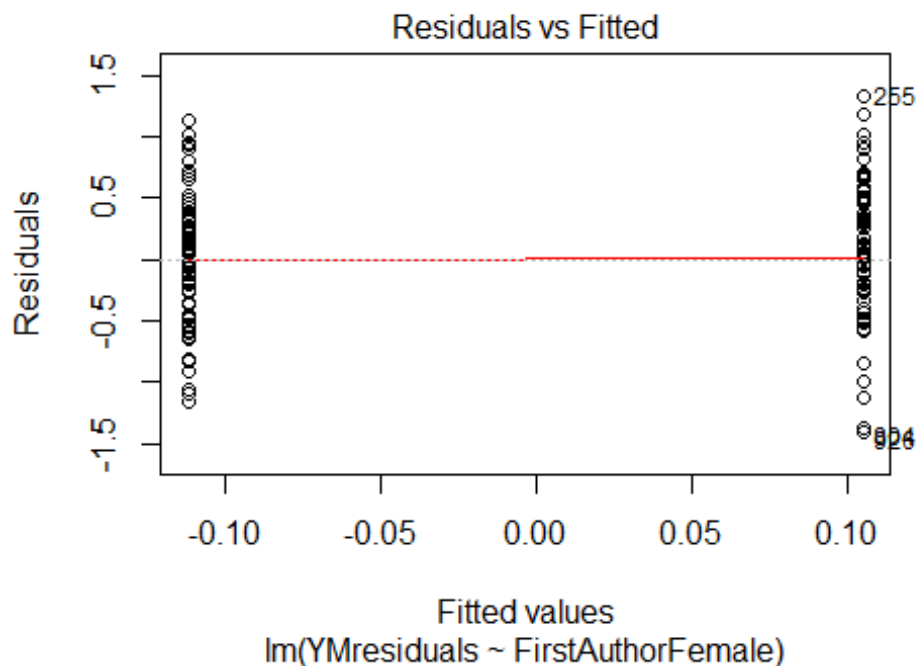
```
## 23 27
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 5 0 2 4 2 4 5 3 11 8 9 8 16 20
## 2011 2012
## 20 24
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 15, p-value = 0.5
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.36, df = 1, p-value = 0.5
## [1] "Female first author team size 2018 geometric mean: 4.16940525104117"
## [1] "Male first author team size 2018 geometric mean: 3.98249845181447"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

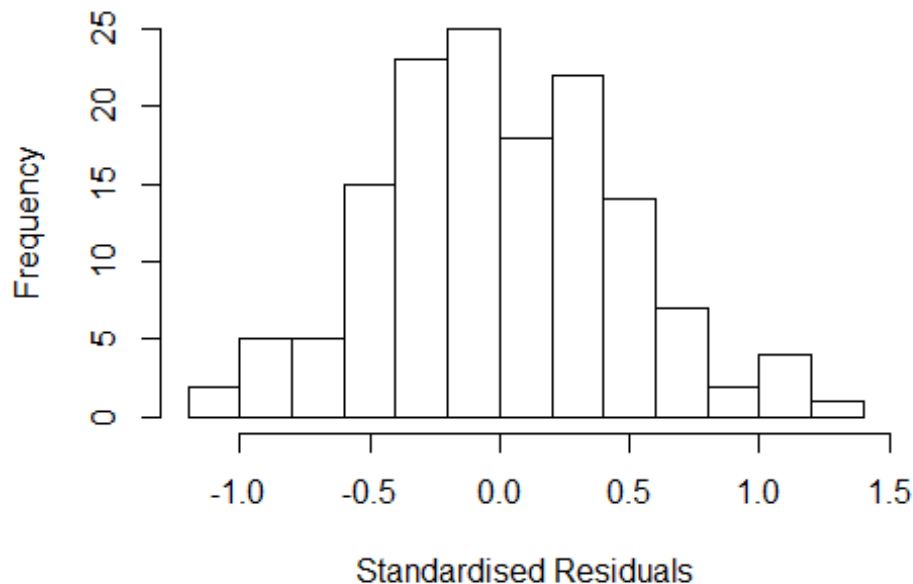
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 23, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.47125171471254"
## [1] "Male last author team size 2018 geometric mean: 4.40931547081958"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 14, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.813 1 1.677
## LastAuthorFemale 1.840 1 1.356
## UniqueAuthors 11.564 4 1.358
## Year 9.064 15 1.076
```

## Residuals from first and last author and team size



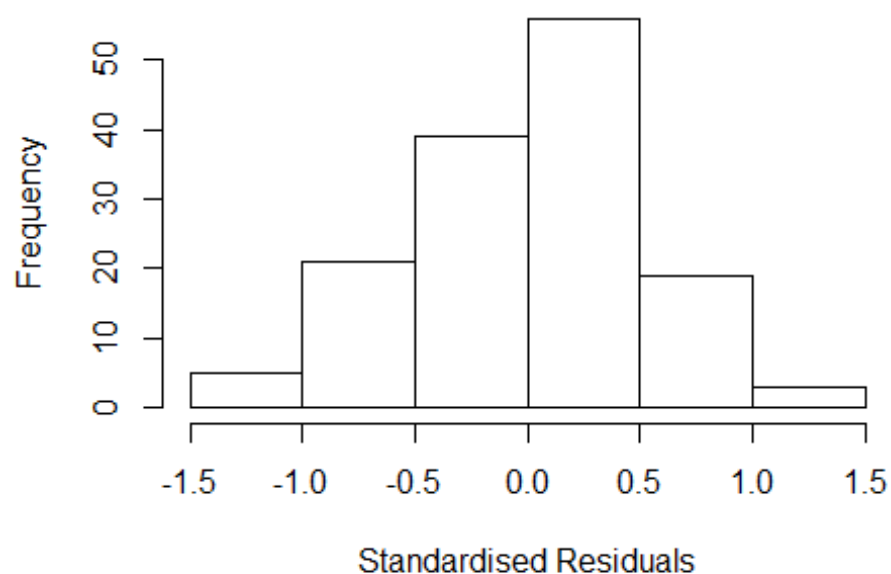
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1487 -0.2972 -0.0342 0.2762 1.3687
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.20950 0.15312 1.37 0.17377
## FirstAuthorFemale1 0.11775 0.11634 1.01 0.31352
## LastAuthorFemale1 0.06976 0.10967 0.64 0.52593
## UniqueAuthors2 0.43412 0.15639 2.78 0.00638 **
## UniqueAuthors3 0.48103 0.17368 2.77 0.00650 **
## UniqueAuthors4 0.53847 0.16587 3.25 0.00151 **
## UniqueAuthors5 0.68915 0.17242 4.00 0.00011 ***
## Year1997 0.74286 0.22692 3.27 0.00138 **
## Year1999 -0.00396 0.70294 -0.01 0.99552
## Year2000 -0.17927 0.48825 -0.37 0.71414
```

```

## Year2001      0.11939      0.31145      0.38  0.70214
## Year2002      0.09006      0.25942      0.35  0.72906
## Year2003      0.45654      0.22583      2.02  0.04542 *
## Year2004      0.58133      0.24993      2.33  0.02169 *
## Year2005      0.31077      0.21545      1.44  0.15178
## Year2006      0.34573      0.21371      1.62  0.10832
## Year2007      0.08556      0.25720      0.33  0.73997
## Year2008      0.54284      0.24678      2.20  0.02973 *
## Year2009      0.42603      0.24462      1.74  0.08412 .
## Year2010      0.49973      0.19777      2.53  0.01280 *
## Year2011      0.38733      0.26601      1.46  0.14796
## Year2012      0.51923      0.20646      2.51  0.01322 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.498
## Multiple R-squared:  0.329, Adjusted R-squared:  0.212
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 135 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.430  0.894  0.966  0.916  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.99e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.043 1 1.429
## LastAuthorFemale 1.412 1 1.188
## Year 2.341 15 1.029

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4891 -0.3369 0.0305 0.3040 1.4915
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2095 0.1523 1.38 0.17151
## FirstAuthorFemale1 0.2434 0.1124 2.17 0.03228 *
## LastAuthorFemale1 0.1333 0.1048 1.27 0.20580
## Year1997 0.9302 0.2442 3.81 0.00022 ***
## Year1999 0.4219 0.6279 0.67 0.50289
## Year2000 0.0636 0.4022 0.16 0.87459
## Year2001 0.3258 0.4934 0.66 0.51029
## Year2002 0.2527 0.2507 1.01 0.31537
## Year2003 0.6418 0.2242 2.86 0.00492 **
## Year2004 0.8102 0.2692 3.01 0.00316 **
## Year2005 0.6497 0.2274 2.86 0.00502 **
## Year2006 0.7637 0.2194 3.48 0.00069 ***
```

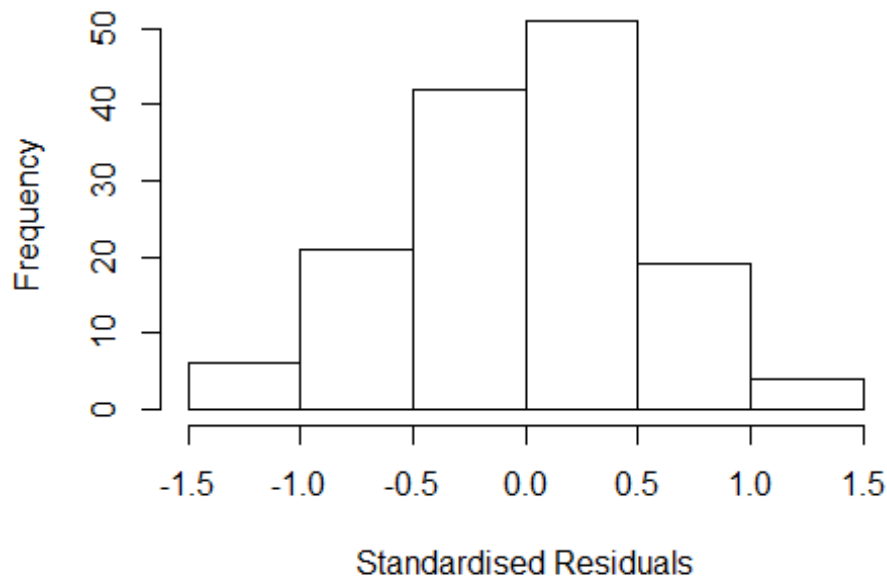


```

## Year2007          0.5235      0.2389      2.19  0.03030 *
## Year2008          0.9223      0.2030      4.54  1.3e-05 ***
## Year2009          0.9029      0.2282      3.96  0.00013 ***
## Year2010          0.8189      0.1991      4.11  7.0e-05 ***
## Year2011          0.7322      0.2330      3.14  0.00209 **
## Year2012          0.7866      0.1942      4.05  8.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.541
## Multiple R-squared:  0.218, Adjusted R-squared:  0.111
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 132 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.427  0.883   0.956   0.913   0.990   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      6.99e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.822 1      1.35
## Year              1.822 15      1.02

```

## Residuals from first author



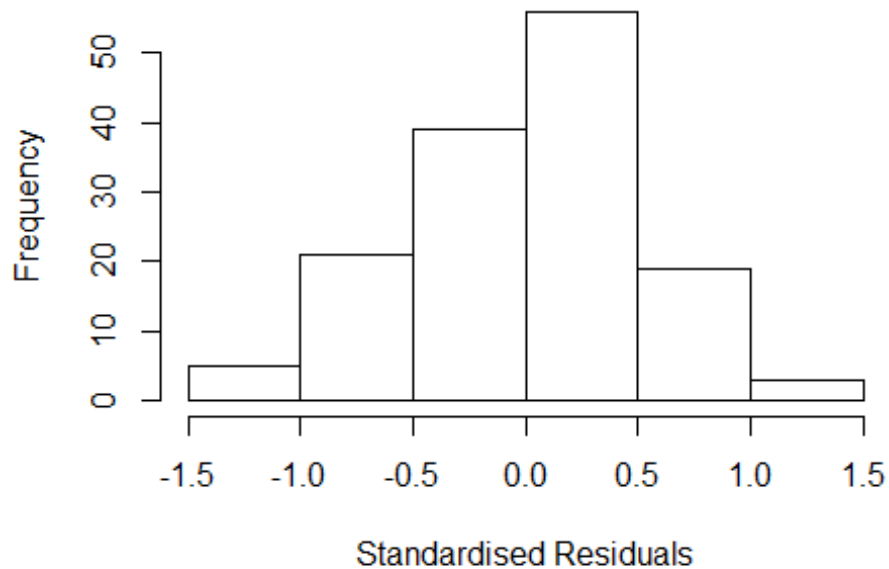
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4060 -0.3622 0.0189 0.3397 1.4071
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.210 0.152 1.37 0.17177
## FirstAuthorFemale1 0.278 0.107 2.60 0.01030 *
## Year1997 0.945 0.234 4.05 9.0e-05 ***
## Year1999 0.488 0.760 0.64 0.52133
## Year2000 0.114 0.380 0.30 0.76520
## Year2001 0.309 0.474 0.65 0.51645
## Year2002 0.311 0.256 1.21 0.22720
## Year2003 0.656 0.223 2.94 0.00388 **
## Year2004 0.843 0.251 3.36 0.00102 **
## Year2005 0.682 0.223 3.06 0.00267 **
## Year2006 0.812 0.223 3.65 0.00039 ***
## Year2007 0.546 0.237 2.31 0.02273 *
```

```

## Year2008          0.957      0.203      4.72  6.1e-06 ***
## Year2009          0.919      0.234      3.93  0.00014 ***
## Year2010          0.856      0.200      4.28  3.7e-05 ***
## Year2011          0.764      0.234      3.26  0.00143 **
## Year2012          0.817      0.197      4.15  6.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.535
## Multiple R-squared:  0.205, Adjusted R-squared:  0.104
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 128 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.468  0.886  0.956  0.908  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.99e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.26 1          1.123
## Year            1.26 15          1.008

```

## Residuals from last author



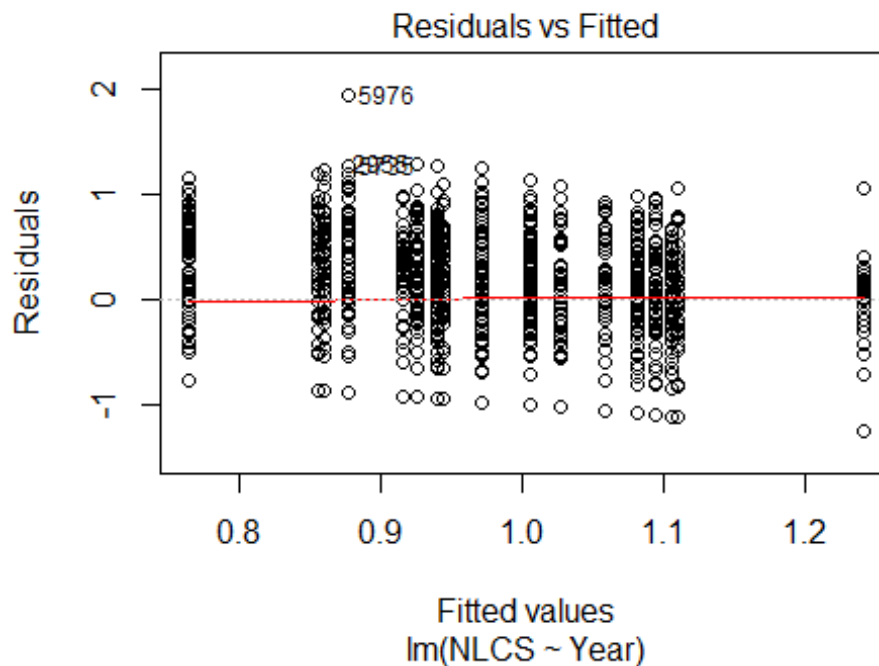
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4330 -0.3381 0.0648 0.3345 1.8290
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.2095 0.1523 1.38 0.17143
## LastAuthorFemale1 0.1930 0.0984 1.96 0.05199 .
## Year1997 0.9585 0.2239 4.28 3.6e-05 ***
## Year1999 0.3920 0.5799 0.68 0.50032
## Year2000 -0.0305 0.3456 -0.09 0.92983
## Year2001 0.4475 0.6690 0.67 0.50478
## Year2002 0.2718 0.2857 0.95 0.34330
## Year2003 0.7360 0.2248 3.27 0.00137 **
## Year2004 0.8717 0.2371 3.68 0.00035 ***
## Year2005 0.7886 0.2304 3.42 0.00084 ***
## Year2006 0.8007 0.2062 3.88 0.00017 ***
## Year2007 0.6561 0.2209 2.97 0.00357 **
```

```

## Year2008          1.0717      0.2046      5.24  6.6e-07 ***
## Year2009          1.0305      0.2144      4.81  4.3e-06 ***
## Year2010          0.9490      0.1960      4.84  3.7e-06 ***
## Year2011          0.8333      0.2382      3.50  0.00065 ***
## Year2012          0.8877      0.1872      4.74  5.7e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.194, Adjusted R-squared:  0.0921
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 10 weights are ~= 1. The remaining 133 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.232  0.877  0.958  0.911  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.99e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 143"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3500"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 272 296 327 317 354 357 268 211 241 276 295 416 327 303 235
## 2011 2012
## 215 262
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 38 50 56 59 58 56 74 59 73 95 123 148 141 97 67
## 2011 2012
## 75 120

```

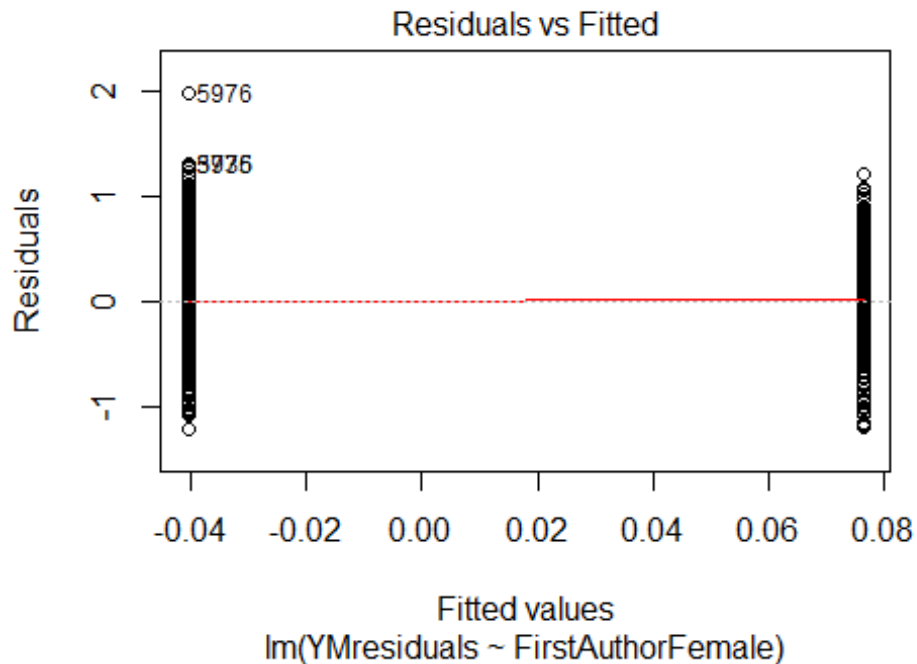
```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   40   53   50   47   36   63   49   64   75   96  119  125   78   51
## 2011 2012
##   52   88
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 19, df = 16, p-value = 0.3
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.7, df = 1, p-value = 0.1

## [1] "Female first author team size 2018 geometric mean: 2.95772814921793"
## [1] "Male first author team size 2018 geometric mean: 2.11477463014356"

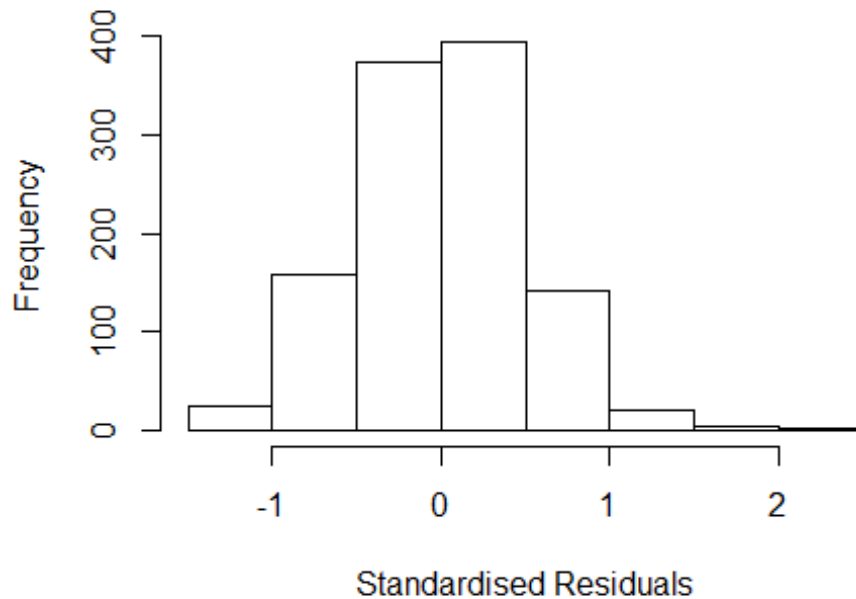
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 1400, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.4762811427719"
## [1] "Male last author team size 2018 geometric mean: 2.47284115020728"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 840, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.214	1	1.102
LastAuthorFemale	1.215	1	1.102
UniqueAuthors	1.619	4	1.062
Year	1.704	16	1.017

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.28983 -0.32249 0.00107 0.32250 2.32389
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8862 0.0885 10.01 < 2e-16 ***
## FirstAuthorFemale1 0.0847 0.0331 2.56 0.01070 *
## LastAuthorFemale1 -0.0287 0.0371 -0.77 0.43968
## UniqueAuthors2 0.4036 0.0504 8.01 2.8e-15 ***
## UniqueAuthors3 0.4717 0.0472 10.00 < 2e-16 ***
## UniqueAuthors4 0.5233 0.0491 10.66 < 2e-16 ***
## UniqueAuthors5 0.6553 0.0492 13.32 < 2e-16 ***
## Year1997 -0.3163 0.1159 -2.73 0.00647 **
## Year1998 -0.0957 0.0948 -1.01 0.31300
## Year1999 -0.2446 0.1125 -2.18 0.02981 *
```

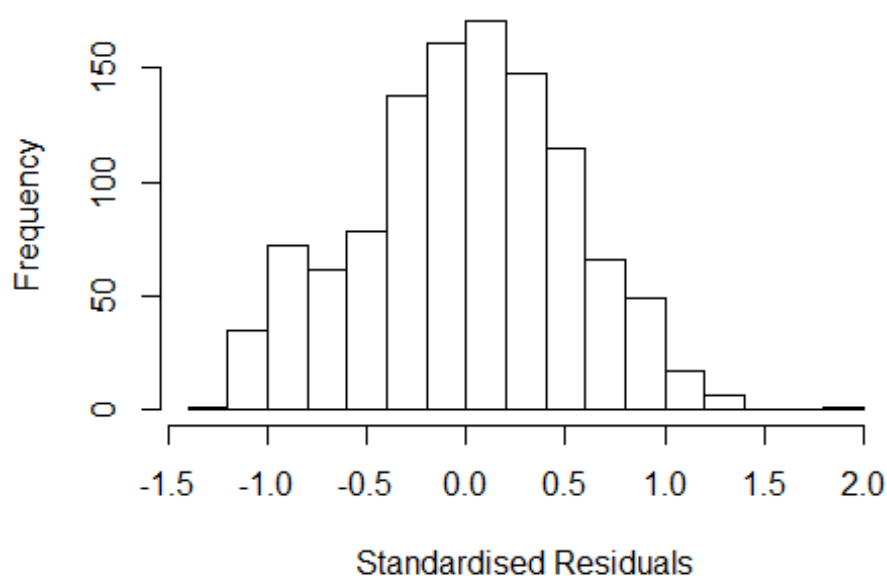


```

## Year2000          -0.2272      0.1108    -2.05   0.04061 *
## Year2001          -0.2310      0.1331    -1.74   0.08293 .
## Year2002          -0.2411      0.0996    -2.42   0.01568 *
## Year2003          -0.2965      0.0992    -2.99   0.00287 **
## Year2004          -0.3513      0.0976    -3.60   0.00034 ***
## Year2005          -0.2990      0.0954    -3.13   0.00177 **
## Year2006          -0.4588      0.0987    -4.65   3.7e-06 ***
## Year2007          -0.3981      0.0903    -4.41   1.2e-05 ***
## Year2008          -0.2925      0.0933    -3.13   0.00177 **
## Year2009          -0.3581      0.0994    -3.60   0.00033 ***
## Year2010          -0.2609      0.1008    -2.59   0.00976 **
## Year2011          -0.3142      0.1100    -2.86   0.00436 **
## Year2012          -0.3921      0.0987    -3.97   7.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.486
## Multiple R-squared:  0.22,   Adjusted R-squared:  0.205
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## observation 1119 is an outlier with |weight| = 0 ( < 8.9e-05);
## 92 weights are ~= 1. The remaining 1026 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.183  0.870  0.951  0.907  0.985  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          8.94e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500          50          2          1          1000          200
## trace.lev mts compute.rd
##           0          1000          0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.152 1          1.073
## LastAuthorFemale 1.162 1          1.078
## Year 1.150 16          1.004

```

## Residuals from first and last author



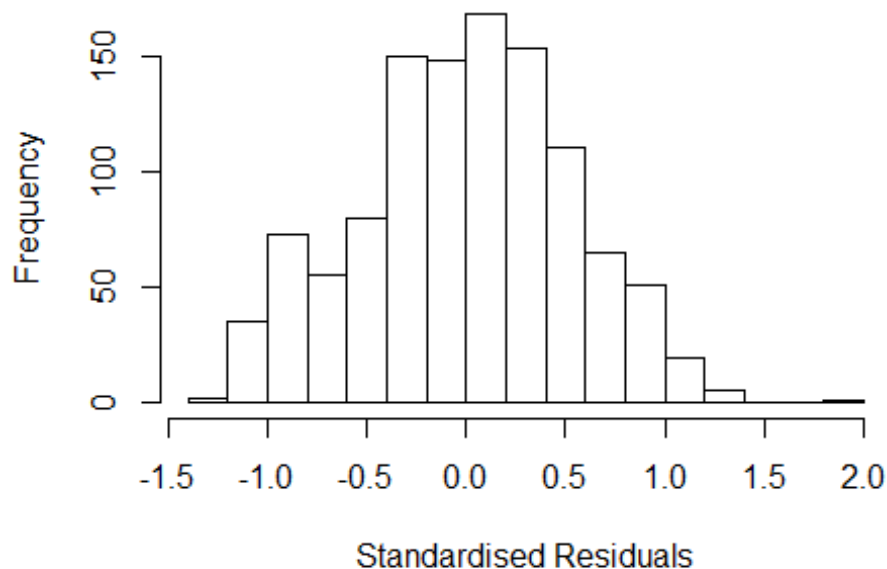
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2453 -0.3726  0.0213  0.3596  1.9252
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2453     0.0678   18.37 < 2e-16 ***
## FirstAuthorFemale1  0.1402     0.0359    3.91 1.0e-04 ***
## LastAuthorFemale1 -0.0895     0.0410   -2.18 0.02932 *
## Year1997          -0.2788     0.1199   -2.32 0.02030 *
## Year1998          -0.1353     0.0891   -1.52 0.12888
## Year1999          -0.3260     0.1055   -3.09 0.00204 **
## Year2000          -0.2097     0.1067   -1.97 0.04962 *
## Year2001          -0.2402     0.1414   -1.70 0.08969 .
## Year2002          -0.1992     0.0943   -2.11 0.03490 *
## Year2003          -0.2004     0.0906   -2.21 0.02721 *
## Year2004          -0.3967     0.1011   -3.93 9.2e-05 ***
## Year2005          -0.3171     0.0928   -3.42 0.00065 ***
```

```

## Year2006          -0.5771      0.0879    -6.57  7.9e-11 ***
## Year2007          -0.3028      0.0811    -3.73  0.00020 ***
## Year2008          -0.2248      0.0859    -2.62  0.00902 **
## Year2009          -0.3563      0.0951    -3.75  0.00019 ***
## Year2010          -0.2565      0.1026    -2.50  0.01259 *
## Year2011          -0.2818      0.1065    -2.65  0.00826 **
## Year2012          -0.3525      0.0935    -3.77  0.00017 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.53
## Multiple R-squared:  0.058, Adjusted R-squared:  0.0426
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 93 weights are ~= 1. The remaining 1026 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.160  0.861  0.950  0.907  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.081 1          1.039
## Year              1.081 16          1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2281 -0.3684  0.0187  0.3597  1.9380
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2281     0.0673   18.24 < 2e-16 ***
## FirstAuthorFemale1 0.1111     0.0353    3.15  0.00166 **
## Year1997         -0.2647     0.1199   -2.21  0.02746 *
## Year1998         -0.1252     0.0895   -1.40  0.16217
## Year1999         -0.3382     0.1053   -3.21  0.00136 **
## Year2000         -0.2013     0.1053   -1.91  0.05618 .
## Year2001         -0.2341     0.1391   -1.68  0.09268 .
## Year2002         -0.1912     0.0940   -2.03  0.04213 *
## Year2003         -0.1924     0.0914   -2.10  0.03552 *
## Year2004         -0.3933     0.1018   -3.86  0.00012 ***
## Year2005         -0.3110     0.0932   -3.34  0.00087 ***
## Year2006         -0.5719     0.0882   -6.49  1.3e-10 ***
```

```

## Year2007          -0.2919      0.0812   -3.59  0.00034 ***
## Year2008          -0.2204      0.0857   -2.57  0.01028 *
## Year2009          -0.3518      0.0958   -3.67  0.00025 ***
## Year2010          -0.2521      0.1034   -2.44  0.01488 *
## Year2011          -0.2775      0.1056   -2.63  0.00870 **
## Year2012          -0.3481      0.0938   -3.71  0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.0541, Adjusted R-squared:  0.0395
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 102 weights are ~= 1. The remaining 1017 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.159  0.858  0.951  0.907  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.088 1      1.043
## Year      1.088 16      1.003

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.2693 -0.3576  0.0206  0.3705  1.8961

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2693    0.0665   19.09 < 2e-16 ***
## LastAuthorFemale1 -0.0329    0.0393   -0.84  0.40151
## Year1997          -0.2648    0.1226   -2.16  0.03090 *
## Year1998          -0.1213    0.0887   -1.37  0.17196
## Year1999          -0.3038    0.1055   -2.88  0.00407 **
## Year2000          -0.1898    0.1059   -1.79  0.07332 .
## Year2001          -0.2028    0.1432   -1.42  0.15697
## Year2002          -0.1967    0.0943   -2.09  0.03724 *
## Year2003          -0.1775    0.0896   -1.98  0.04771 *
## Year2004          -0.3899    0.0998   -3.91  9.9e-05 ***
## Year2005          -0.3037    0.0918   -3.31  0.00097 ***
## Year2006          -0.5653    0.0878   -6.44  1.8e-10 ***
## Year2007          -0.2759    0.0802   -3.44  0.00061 ***
## Year2008          -0.2190    0.0851   -2.57  0.01018 *
## Year2009          -0.3474    0.0953   -3.65  0.00028 ***
## Year2010          -0.2536    0.1031   -2.46  0.01410 *
## Year2011          -0.2799    0.1063   -2.63  0.00856 **
## Year2012          -0.3474    0.0936   -3.71  0.00022 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.538
## Multiple R-squared:  0.0455, Adjusted R-squared:  0.0308
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 88 weights are ~= 1. The remaining 1031 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.188  0.857  0.948  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.94e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000        200
##   trace.lev    mts    compute.rd
##      0        1000         0
##      psi          subsampling          cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1119"
## [1] ""

```

```

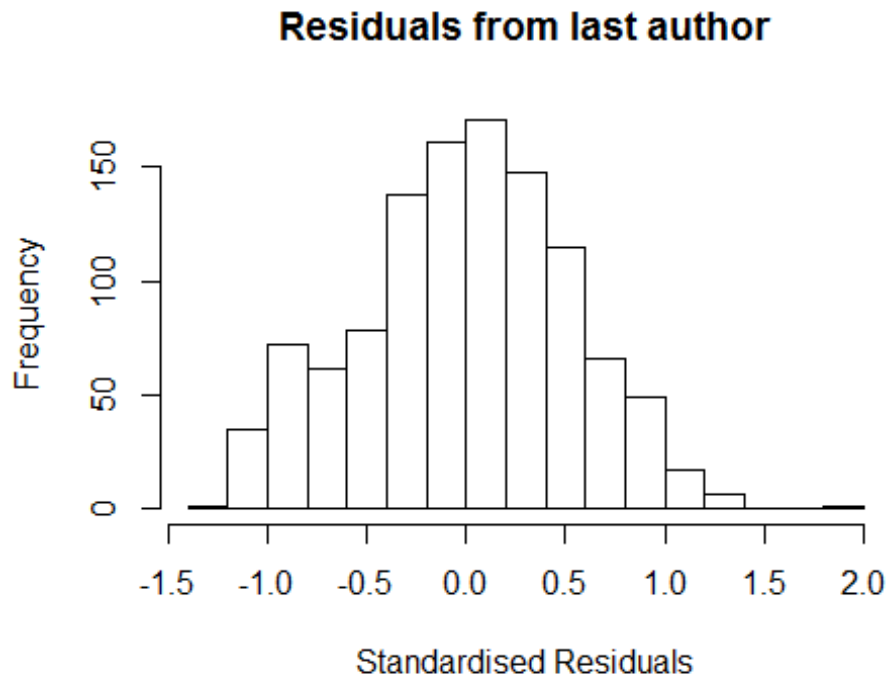
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3501"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2005 2006 2007 2008 2009 2010 2011 2012
## 10 7 6 2 3 8 10 12 7 2 8 13
##
## 1996 1997 1998 1999 2005 2006 2007 2008 2009 2010 2011 2012
## 0 1 0 0 0 0 0 0 0 1 0 5
##
## 1996 1997 1998 1999 2005 2006 2007 2008 2009 2010 2011 2012
## 0 1 0 0 0 0 0 0 0 1 0 5
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 6"
## [1] "Male first author team size 2018 geometric mean: 2.71441761659491"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2"
## [1] "Male last author team size 2018 geometric mean: 3.91486764116886"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



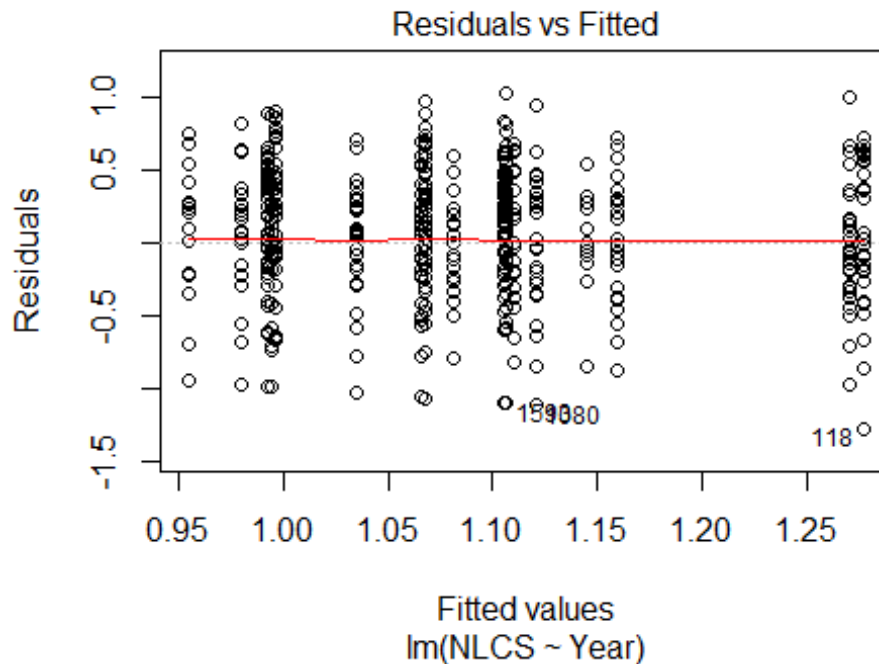
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0.5, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## [1] "Regression 2: First author gender, Last author gender, Year as
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 7"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3503"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
## [1] "gendered] [check that these decrease]"
##
## 2003 2004 2007 2011
##    1    4    2    1
##
## 2003 2004 2007 2011
##    0    3    0    1
```



```

##
## 2003 2004 2007 2011
##    0    1    0    1
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: NaN"
## [1] "Male first author team size 2018 geometric mean: NaN"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 2"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3504"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    87   94   87  112  126  109  114   88   74   58   79   90  122  147  141
## 2011 2012
##   125  146
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    18   32   16   20   29   27   33   38   26   28   27   36   60   72   75
## 2011 2012
##    65   69
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    17   28   12   12   27   21   29   24   14   18   19   34   54   67   67
## 2011 2012
##    53   56
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6

```



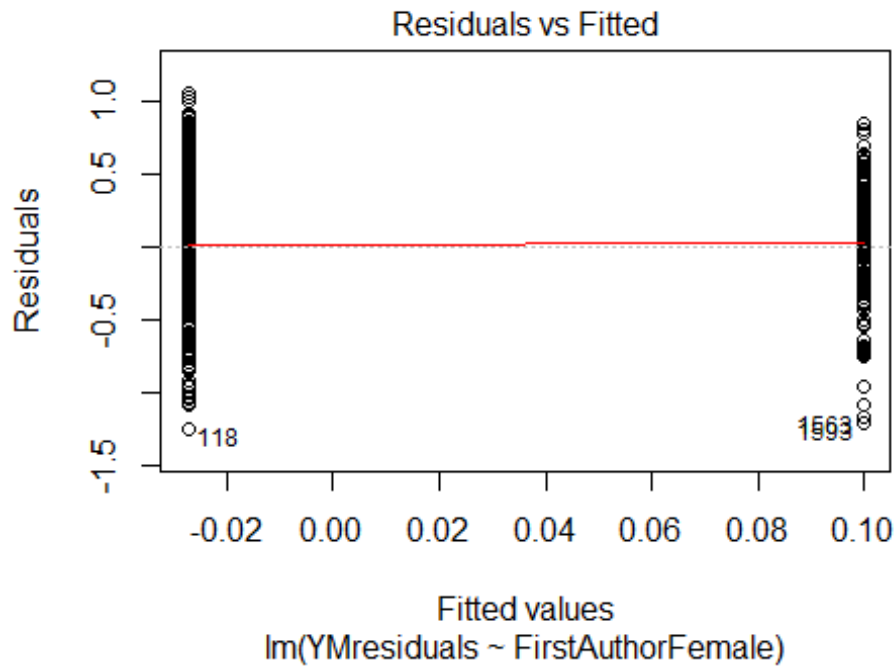
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.57, df = 1, p-value = 0.4

## [1] "Female first author team size 2018 geometric mean: 2.95796267181364"
## [1] "Male first author team size 2018 geometric mean: 3.57492468726009"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

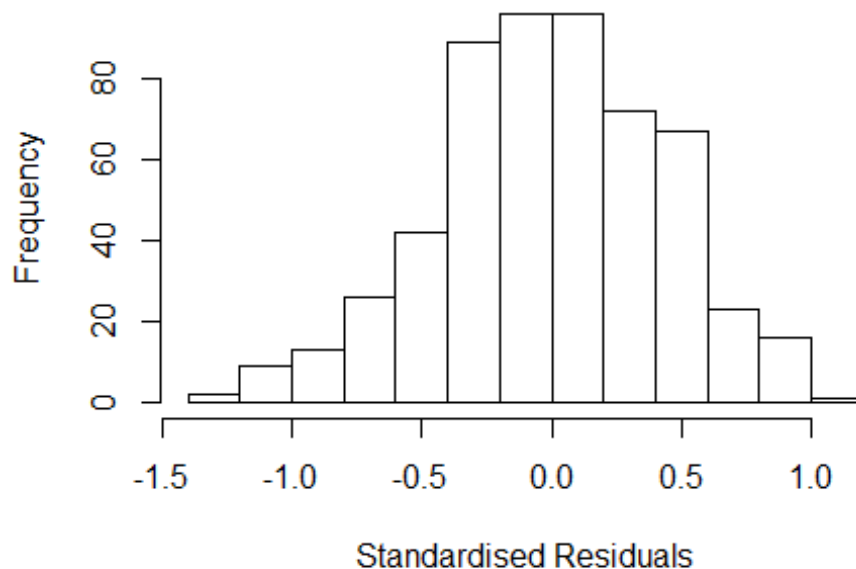
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 100, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.68179283050743"
## [1] "Male last author team size 2018 geometric mean: 3.64630566281187"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 21, p-value = 0.04
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.115 1          1.056
## LastAuthorFemale  1.139 1          1.067
## UniqueAuthors    1.663 4          1.066
## Year              1.796 16         1.018
```

## Residuals from first and last author and team size



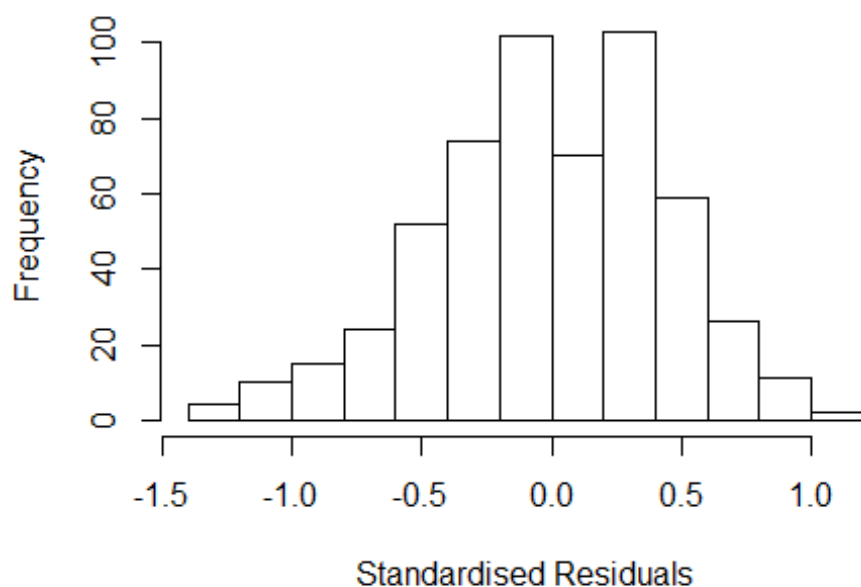
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3597 -0.2879 -0.0042 0.3077 1.0720
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.80995 0.15050 5.38 1.1e-07 ***
## FirstAuthorFemale1 0.12330 0.04366 2.82 0.00492 **
## LastAuthorFemale1 0.09125 0.05502 1.66 0.09783 .
## UniqueAuthors2 0.13154 0.10812 1.22 0.22428
## UniqueAuthors3 0.20162 0.10457 1.93 0.05438 .
## UniqueAuthors4 0.21186 0.10450 2.03 0.04313 *
## UniqueAuthors5 0.37702 0.10493 3.59 0.00036 ***
## Year1997 0.26509 0.17035 1.56 0.12028
## Year1998 0.25722 0.15360 1.67 0.09459 .
## Year1999 0.01813 0.16211 0.11 0.91102
```

```

## Year2000      -0.03311      0.15575      -0.21      0.83173
## Year2001      0.23496      0.17119      1.37      0.17048
## Year2002      0.03477      0.14720      0.24      0.81334
## Year2003      0.05862      0.15276      0.38      0.70135
## Year2004     -0.01356      0.19986     -0.07      0.94591
## Year2005     -0.10143      0.16614     -0.61      0.54177
## Year2006      0.05109      0.17645      0.29      0.77226
## Year2007      0.04739      0.15139      0.31      0.75439
## Year2008      0.06803      0.14545      0.47      0.64015
## Year2009      0.00531      0.14634      0.04      0.97109
## Year2010      0.04947      0.14226      0.35      0.72820
## Year2011     -0.07265      0.14599     -0.50      0.61893
## Year2012     -0.00698      0.14400     -0.05      0.96135
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.433
## Multiple R-squared:  0.112, Adjusted R-squared:  0.0752
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.302  0.881  0.951  0.908  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          1.81e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.136 1      1.066
## LastAuthorFemale  1.113 1      1.055
## Year              1.150 16      1.004

```

## Residuals from first and last author



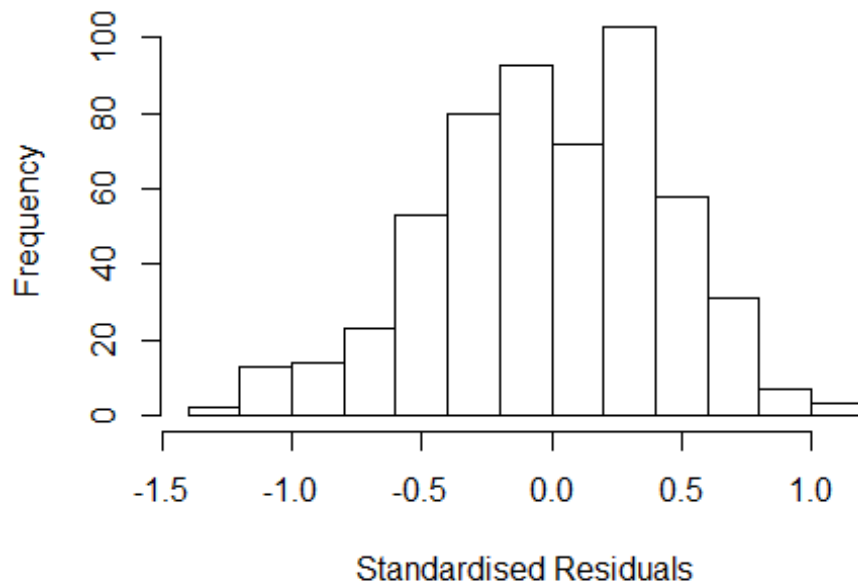
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.27228 -0.30482 -0.00809 0.31735 1.05317
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.959992 0.136344 7.04 5.9e-12 ***
## FirstAuthorFemale1 0.138183 0.044600 3.10 0.002 **
## LastAuthorFemale1 0.100900 0.054865 1.84 0.066 .
## Year1997 0.312290 0.175834 1.78 0.076 .
## Year1998 0.239629 0.149516 1.60 0.110
## Year1999 0.137136 0.164488 0.83 0.405
## Year2000 0.019901 0.158921 0.13 0.900
## Year2001 0.272628 0.172786 1.58 0.115
## Year2002 0.090402 0.154973 0.58 0.560
## Year2003 0.100085 0.158840 0.63 0.529
## Year2004 0.050903 0.187296 0.27 0.786
## Year2005 -0.056345 0.174416 -0.32 0.747
```

```

## Year2006          0.154538    0.180516    0.86    0.392
## Year2007          0.142156    0.156784    0.91    0.365
## Year2008          0.145371    0.147549    0.99    0.325
## Year2009          0.029971    0.150297    0.20    0.842
## Year2010          0.119836    0.147489    0.81    0.417
## Year2011          0.000634    0.152697    0.00    0.997
## Year2012          0.095513    0.149361    0.64    0.523
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.0604, Adjusted R-squared:  0.0287
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 512 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.386  0.885   0.949   0.907   0.982   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.077 1         1.038
## Year              1.077 16         1.002

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27996 -0.28195 -0.00984  0.32168  1.04277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.96507    0.13273   7.27 1.3e-12 ***
## FirstAuthorFemale1 0.15603    0.04388   3.56 0.00041 ***
## Year1997        0.31489    0.17150   1.84 0.06690 .
## Year1998        0.24057    0.14631   1.64 0.10071
## Year1999        0.13569    0.16157   0.84 0.40137
## Year2000        0.02023    0.15561   0.13 0.89662
## Year2001        0.27583    0.16921   1.63 0.10368
## Year2002        0.09365    0.15172   0.62 0.53733
## Year2003        0.10550    0.15645   0.67 0.50038
## Year2004        0.05325    0.18582   0.29 0.77457
## Year2005       -0.04519    0.17102  -0.26 0.79168
## Year2006        0.16583    0.17839   0.93 0.35299
```

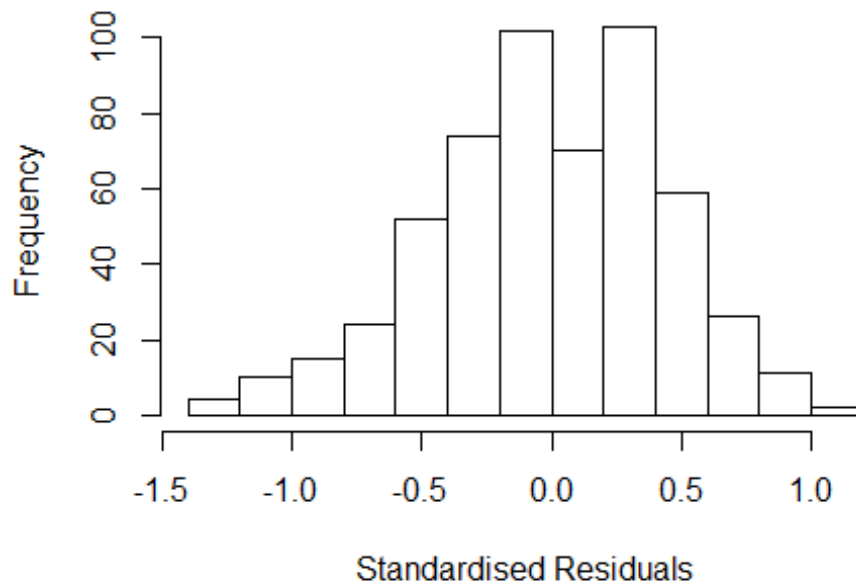


```

## Year2007          0.15604      0.15421      1.01  0.31208
## Year2008          0.15525      0.14462      1.07  0.28352
## Year2009          0.03421      0.14716      0.23  0.81626
## Year2010          0.12516      0.14484      0.86  0.38790
## Year2011          0.00211      0.14966      0.01  0.98877
## Year2012          0.09777      0.14609      0.67  0.50365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.449
## Multiple R-squared:  0.0543, Adjusted R-squared:  0.0242
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 42 weights are ~= 1. The remaining 510 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.397  0.886  0.950  0.909  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.076 1      1.037
## Year              1.076 16      1.002

```

## Residuals from last author



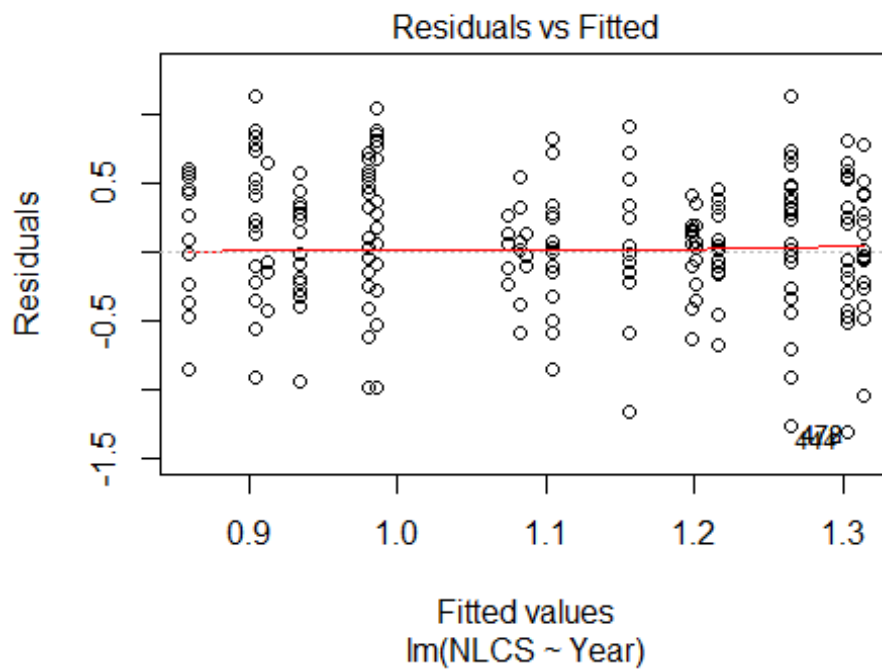
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.29884 -0.29064 -0.00376  0.31168  1.03380
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.9659     0.1411   6.84 2.1e-11 ***
## LastAuthorFemale1  0.1355     0.0546   2.48  0.013 *
## Year1997          0.3330     0.1846   1.80  0.072 .
## Year1998          0.2542     0.1543   1.65  0.100
## Year1999          0.1761     0.1653   1.07  0.287
## Year2000          0.0349     0.1619   0.22  0.830
## Year2001          0.2867     0.1789   1.60  0.110
## Year2002          0.1060     0.1584   0.67  0.504
## Year2003          0.1399     0.1637   0.85  0.393
## Year2004          0.0524     0.1936   0.27  0.787
## Year2005         -0.0336     0.1822  -0.18  0.854
## Year2006          0.1521     0.1839   0.83  0.409
```

```

## Year2007          0.1639      0.1597      1.03      0.305
## Year2008          0.1721      0.1518      1.13      0.257
## Year2009          0.0611      0.1541      0.40      0.692
## Year2010          0.1333      0.1515      0.88      0.379
## Year2011          0.0259      0.1569      0.17      0.869
## Year2012          0.1228      0.1538      0.80      0.425
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.441
## Multiple R-squared:  0.0457, Adjusted R-squared:  0.0153
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 47 weights are ~= 1. The remaining 505 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.367  0.872  0.948  0.904  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.81e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 552"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3505"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   19   13   25   18   41   33   39   44   38   42   38   36   35   40   41
## 2011 2012
##   28   46
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    4    7    7    6   17   14   11   17    3   15   14   22   26   25   26
## 2011 2012

```

```
## 17 29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 3 6 7 4 15 13 9 14 3 14 10 20 24 23 24
## 2011 2012
## 16 26
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



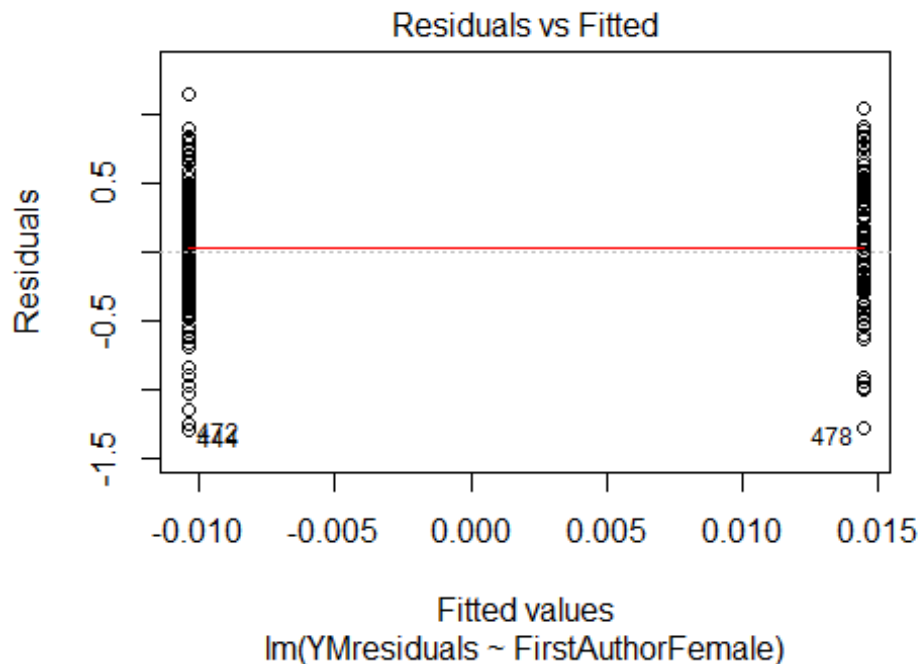
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.095, df = 1, p-value = 0.8

## [1] "Female first author team size 2018 geometric mean: 2.72260464377875"
## [1] "Male first author team size 2018 geometric mean: 2.51984209978975"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 16, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.87037777576898"
## [1] "Male last author team size 2018 geometric mean: 2.47939698673123"

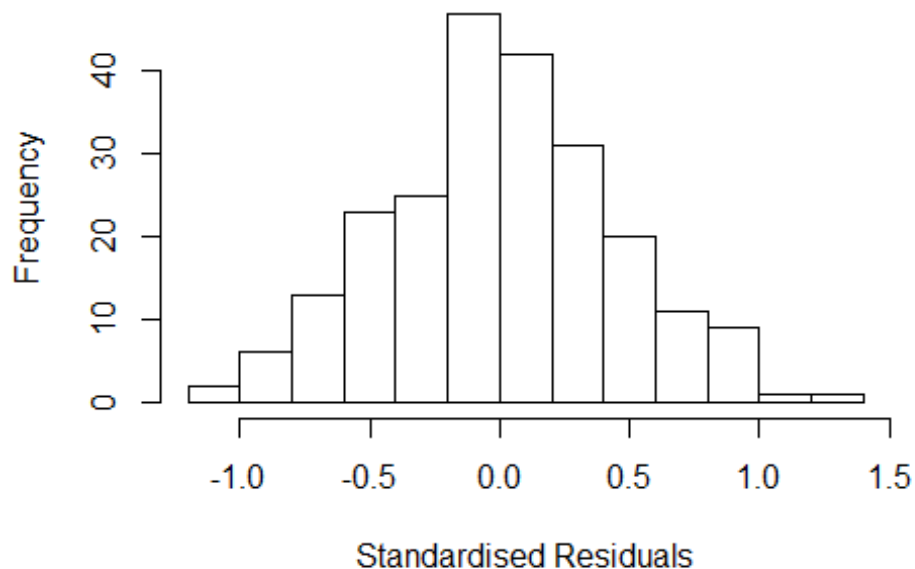
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 31, p-value = 0.8
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.407	1	1.186
LastAuthorFemale	1.649	1	1.284
UniqueAuthors	4.412	4	1.204
Year	6.739	16	1.061

## Residuals from first and last author and team size



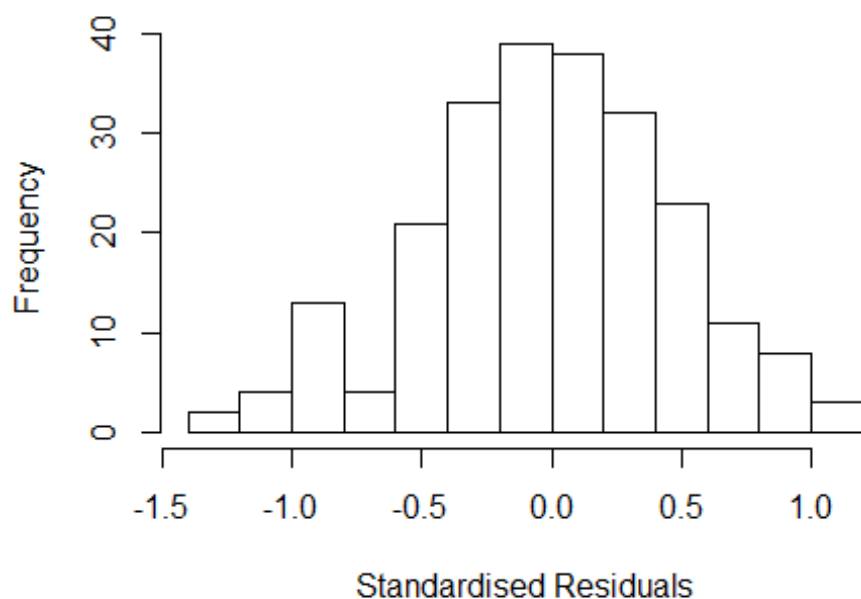
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.12479 -0.25962 -0.00205 0.28711 1.27580
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.61640 0.14190 4.34 2.2e-05 ***
## FirstAuthorFemale1 -0.11598 0.07013 -1.65 0.0997 .
## LastAuthorFemale1 0.07694 0.09284 0.83 0.4082
## UniqueAuthors2 0.31447 0.13448 2.34 0.0203 *
## UniqueAuthors3 0.49584 0.12440 3.99 9.3e-05 ***
## UniqueAuthors4 0.48173 0.14793 3.26 0.0013 **
## UniqueAuthors5 0.58934 0.13481 4.37 1.9e-05 ***
## Year1997 0.24430 0.15510 1.58 0.1168
## Year1998 0.08995 0.18857 0.48 0.6339
## Year1999 0.08503 0.20030 0.42 0.6716
```

```

## Year2000          0.31269    0.17016    1.84    0.0675 .
## Year2001          0.03980    0.20675    0.19    0.8475
## Year2002          0.14782    0.16952    0.87    0.3842
## Year2003         -0.07889    0.16665   -0.47    0.6364
## Year2004          0.00188    0.19477    0.01    0.9923
## Year2005         -0.12691    0.18412   -0.69    0.4914
## Year2006          0.16921    0.15549    1.09    0.2778
## Year2007          0.41999    0.16403    2.56    0.0112 *
## Year2008          0.34915    0.21957    1.59    0.1133
## Year2009         -0.07619    0.18101   -0.42    0.6743
## Year2010          0.06570    0.16698    0.39    0.6944
## Year2011          0.13121    0.19057    0.69    0.4919
## Year2012         -0.01175    0.18229   -0.06    0.9486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.458
## Multiple R-squared:  0.237, Adjusted R-squared:  0.156
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 14 weights are ~= 1. The remaining 217 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.418  0.880  0.961  0.914  0.989  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.33e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.586 1      1.259
## LastAuthorFemale  1.807 1      1.344
## Year              2.635 16      1.031

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.336410 -0.306495 -0.000167 0.294950 1.076885
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7076 0.1194 5.92 1.3e-08 ***
## FirstAuthorFemale1 -0.0546 0.0744 -0.73 0.46360
## LastAuthorFemale1 0.0399 0.1062 0.38 0.70713
## Year1997 0.4880 0.1514 3.22 0.00147 **
## Year1998 0.3956 0.1714 2.31 0.02194 *
## Year1999 0.4232 0.1454 2.91 0.00400 **
## Year2000 0.5942 0.1559 3.81 0.00018 ***
## Year2001 0.3830 0.1835 2.09 0.03805 *
## Year2002 0.5610 0.1360 4.13 5.3e-05 ***
## Year2003 0.2583 0.1493 1.73 0.08499 .
## Year2004 0.3796 0.1317 2.88 0.00436 **
## Year2005 0.1334 0.1650 0.81 0.41967
```

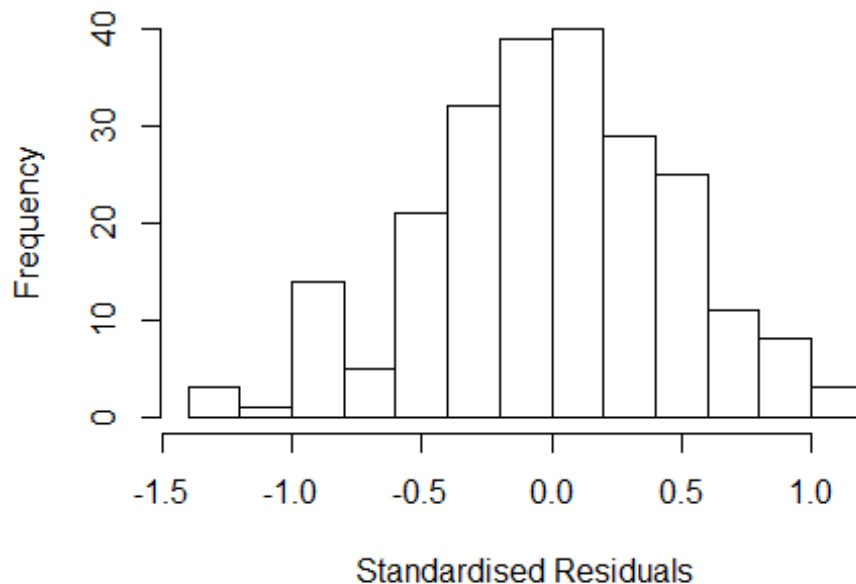


```

## Year2006          0.5055      0.1312      3.85  0.00015 ***
## Year2007          0.6815      0.1381      4.93  1.6e-06 ***
## Year2008          0.6288      0.1756      3.58  0.00043 ***
## Year2009          0.2153      0.1866      1.15  0.24980
## Year2010          0.3261      0.1468      2.22  0.02745 *
## Year2011          0.4812      0.1601      3.01  0.00297 **
## Year2012          0.2612      0.1811      1.44  0.15072
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.469
## Multiple R-squared:  0.121, Adjusted R-squared:  0.0464
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 23 weights are ~= 1. The remaining 208 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.396  0.872  0.950  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.33e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.626 1      1.275
## Year              1.626 16      1.015

```

## Residuals from first author



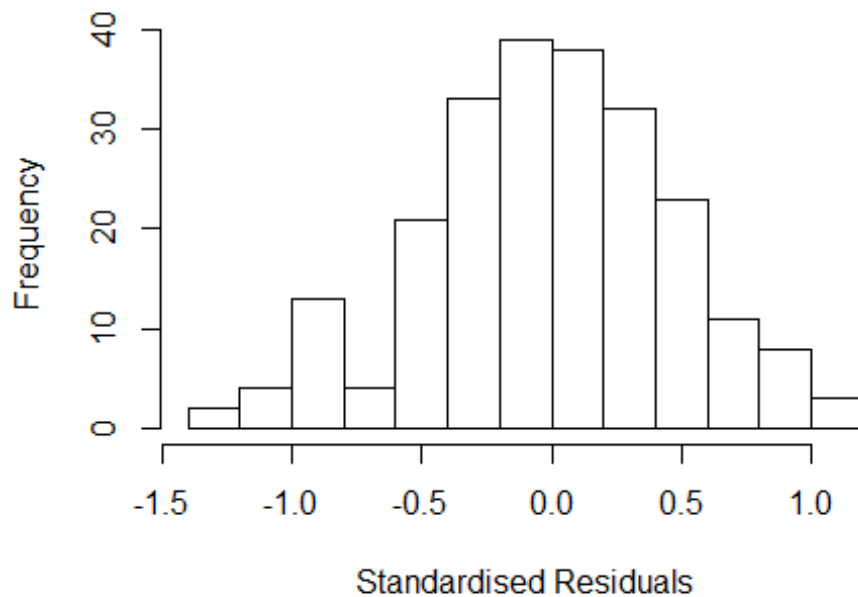
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3419 -0.3157 0.0017 0.3013 1.1084
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.7296 0.0980 7.45 2.3e-12 ***
## FirstAuthorFemale1 -0.0481 0.0757 -0.64 0.52529
## Year1997 0.4628 0.1335 3.47 0.00064 ***
## Year1998 0.3836 0.1707 2.25 0.02567 *
## Year1999 0.3995 0.1272 3.14 0.00193 **
## Year2000 0.5737 0.1415 4.05 7.1e-05 ***
## Year2001 0.3598 0.1694 2.12 0.03484 *
## Year2002 0.5366 0.1160 4.63 6.4e-06 ***
## Year2003 0.2415 0.1381 1.75 0.08176 .
## Year2004 0.3576 0.1126 3.17 0.00172 **
## Year2005 0.1198 0.1594 0.75 0.45290
## Year2006 0.4854 0.1181 4.11 5.6e-05 ***
```

```

## Year2007          0.6654      0.1302      5.11  7.2e-07 ***
## Year2008          0.6123      0.1647      3.72  0.00026 ***
## Year2009          0.2061      0.1824      1.13  0.25966
## Year2010          0.3095      0.1368      2.26  0.02466 *
## Year2011          0.4719      0.1590      2.97  0.00334 **
## Year2012          0.2412      0.1618      1.49  0.13757
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.475
## Multiple R-squared:  0.12,   Adjusted R-squared:  0.0495
## Convergence in 14 IRWLS iterations
##
## Robustness weights:
## 22 weights are ~= 1. The remaining 209 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.405  0.876  0.953  0.905  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.33e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.827 1      1.352
## Year              1.827 16      1.019

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.35031 -0.30975 0.00844 0.30106 1.11099
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6805 0.1155 5.89 1.5e-08 ***
## LastAuthorFemale1 0.0264 0.1094 0.24 0.80967
## Year1997 0.4879 0.1490 3.27 0.00124 **
## Year1998 0.3986 0.1694 2.35 0.01953 *
## Year1999 0.4363 0.1438 3.03 0.00272 **
## Year2000 0.6143 0.1567 3.92 0.00012 ***
## Year2001 0.3976 0.1809 2.20 0.02909 *
## Year2002 0.5690 0.1367 4.16 4.6e-05 ***
## Year2003 0.2790 0.1492 1.87 0.06283 .
## Year2004 0.4066 0.1282 3.17 0.00173 **
## Year2005 0.1377 0.1643 0.84 0.40287
## Year2006 0.5165 0.1334 3.87 0.00014 ***
```

```

## Year2007          0.6782      0.1433      4.73  4.0e-06 ***
## Year2008          0.6434      0.1798      3.58  0.00043 ***
## Year2009          0.2141      0.1896      1.13  0.26006
## Year2010          0.3297      0.1495      2.21  0.02852 *
## Year2011          0.4848      0.1639      2.96  0.00345 **
## Year2012          0.2743      0.1817      1.51  0.13245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.465
## Multiple R-squared:  0.119, Adjusted R-squared:  0.0486
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 211 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.380  0.875  0.950  0.902  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.33e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 231"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3506"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   52   49   31   35   31   39   16   15   29   14   17    9   11    9    8
## 2011 2012
##   19    9
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   10   15    8    8   12   10    5    4    8    5    1    3    7    6    6
## 2011 2012

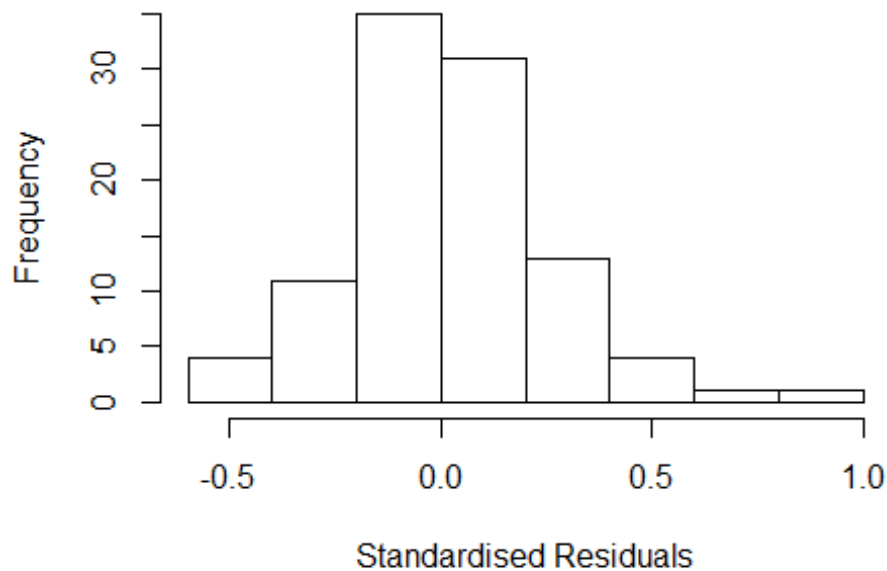
```

```

##      9      6
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##      9     15      5      4      9      8      4      2      6      5      0      2      6      5      6
## 2011 2012
##      8      6
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1"
## [1] "Male first author team size 2018 geometric mean: 4.52701905583787"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.73205080756888"
## [1] "Male last author team size 2018 geometric mean: 5.1924941018511"
##
## Wilcoxon rank sum test
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale  3.104  1      1.762
## LastAuthorFemale   2.951  1      1.718
## UniqueAuthors    110.208  4      1.800
## Year              233.234 15      1.199

```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.46451 -0.13535 -0.00581 0.16009 0.81717
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.11941 0.12648 8.85 2.1e-13 ***
## FirstAuthorFemale1 0.15343 0.08365 1.83 0.070 .
## LastAuthorFemale1 -0.08980 0.11893 -0.76 0.452
## UniqueAuthors2 0.15784 0.13909 1.13 0.260
## UniqueAuthors3 -0.18727 0.22265 -0.84 0.403
## UniqueAuthors4 -0.03790 0.20054 -0.19 0.851
## UniqueAuthors5 0.00185 0.19179 0.01 0.992
## Year1997 -0.14774 0.13897 -1.06 0.291
## Year1998 -0.11500 0.20531 -0.56 0.577
## Year1999 -0.33915 0.19963 -1.70 0.093 .
```

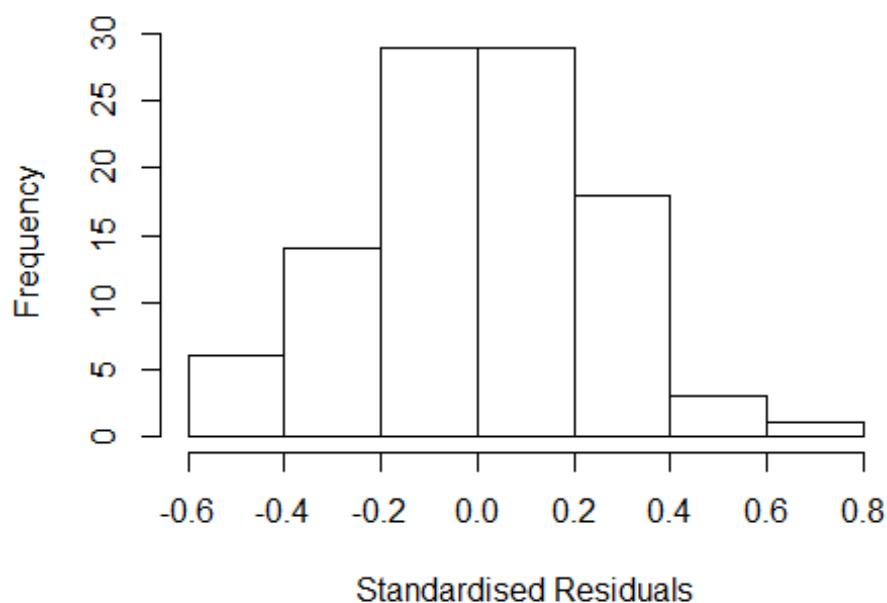
```

## Year2000      -0.10630    0.16822   -0.63    0.529
## Year2001      0.06293    0.18149    0.35    0.730
## Year2002     -0.28027    0.14032   -2.00    0.049 *
## Year2003     -0.04841    0.15923   -0.30    0.762
## Year2004      0.08950    0.12112    0.74    0.462
## Year2005     -0.19159    0.14376   -1.33    0.187
## Year2007      0.39209    0.23936    1.64    0.105
## Year2008     -0.06085    0.24538   -0.25    0.805
## Year2009      0.06371    0.15468    0.41    0.682
## Year2010     -0.06858    0.15195   -0.45    0.653
## Year2011      0.09882    0.20045    0.49    0.623
## Year2012      0.13817    0.16047    0.86    0.392
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.239
## Multiple R-squared:  0.361, Adjusted R-squared:  0.19
## Convergence in 30 IRWLS iterations
##
## Robustness weights:
## 6 weights are ~= 1. The remaining 94 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.217  0.901  0.959  0.913  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.00e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.870 1 1.694
## LastAuthorFemale 1.666 1 1.291
## Year 4.725 15 1.053

```



## Residuals from first and last author



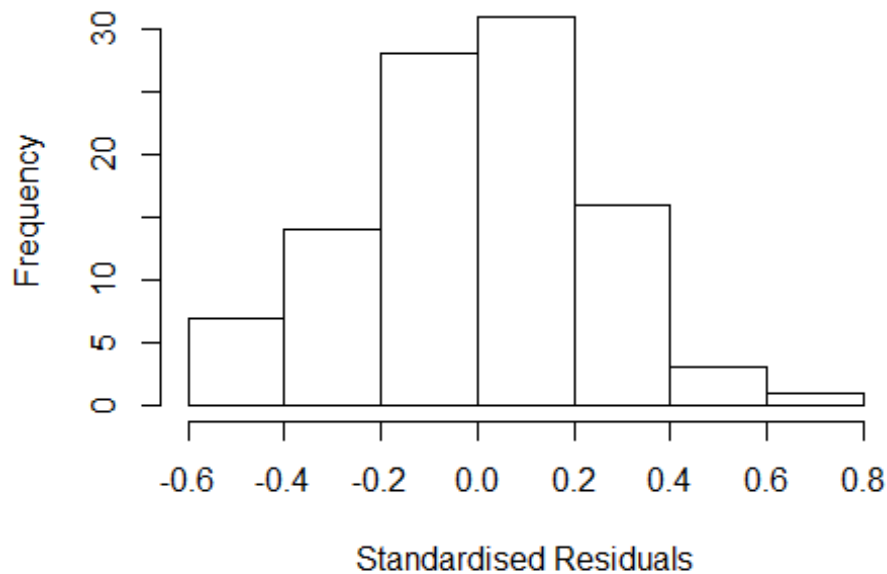
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.57702 -0.17108 0.00575 0.14981 0.61914
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0734 0.1140 9.42 1.1e-14 ***
## FirstAuthorFemale1 0.1655 0.0817 2.03 0.046 *
## LastAuthorFemale1 -0.0830 0.1124 -0.74 0.463
## Year1997 -0.0775 0.1319 -0.59 0.558
## Year1998 -0.0963 0.1718 -0.56 0.577
## Year1999 -0.3689 0.1808 -2.04 0.045 *
## Year2000 -0.1111 0.1465 -0.76 0.450
## Year2001 0.0376 0.1537 0.24 0.807
## Year2002 -0.2095 0.1360 -1.54 0.127
## Year2003 -0.1011 0.1187 -0.85 0.397
## Year2004 0.1886 0.1251 1.51 0.136
## Year2005 -0.1338 0.1337 -1.00 0.320
```

```

## Year2007          0.3394      0.3141      1.08      0.283
## Year2008         -0.0380      0.2418     -0.16      0.875
## Year2009          0.0677      0.1365      0.50      0.621
## Year2010         -0.0297      0.1444     -0.21      0.838
## Year2011          0.1306      0.1646      0.79      0.430
## Year2012          0.1821      0.1416      1.29      0.202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.267
## Multiple R-squared:  0.27,   Adjusted R-squared:  0.118
## Convergence in 16 IRWLS iterations
##
## Robustness weights:
## 7 weights are ~= 1. The remaining 93 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.571  0.887  0.961  0.923  0.990  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.00e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.92 1      1.709
## Year              2.92 15      1.036

```

## Residuals from first author

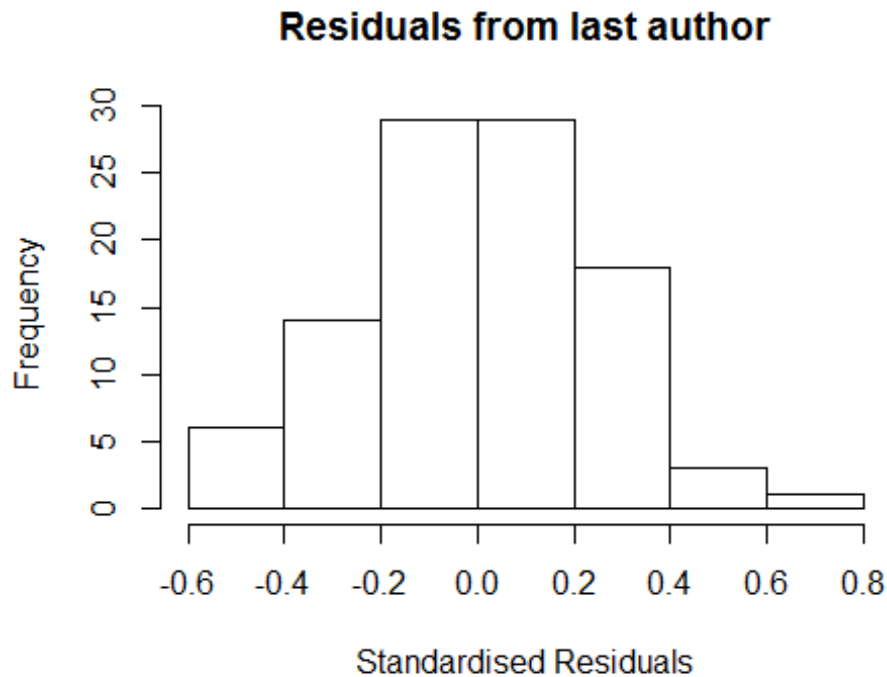


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.5666 -0.1780 0.0198 0.1504 0.6018
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0743 0.1147 9.37 1.2e-14 ***
## FirstAuthorFemale1 0.1575 0.0851 1.85 0.068 .
## Year1997 -0.0786 0.1327 -0.59 0.555
## Year1998 -0.1182 0.1642 -0.72 0.474
## Year1999 -0.3661 0.1838 -1.99 0.050 *
## Year2000 -0.1270 0.1485 -0.86 0.395
## Year2001 0.0262 0.1562 0.17 0.867
## Year2002 -0.2065 0.1360 -1.52 0.133
## Year2003 -0.0981 0.1203 -0.82 0.417
## Year2004 0.1889 0.1255 1.51 0.136
## Year2005 -0.1433 0.1406 -1.02 0.311
## Year2007 0.3424 0.3212 1.07 0.290
```

```

## Year2008          -0.0329      0.2263    -0.15     0.885
## Year2009           0.0509      0.1314     0.39     0.699
## Year2010          -0.0399      0.1484    -0.27     0.789
## Year2011           0.1126      0.1652     0.68     0.497
## Year2012           0.1683      0.1416     1.19     0.238
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.267
## Multiple R-squared:  0.263, Adjusted R-squared:  0.121
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 3 weights are ~= 1. The remaining 97 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.589  0.880  0.964  0.926  0.991  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.00e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.839 1          1.356
## Year             1.839 15          1.021

```



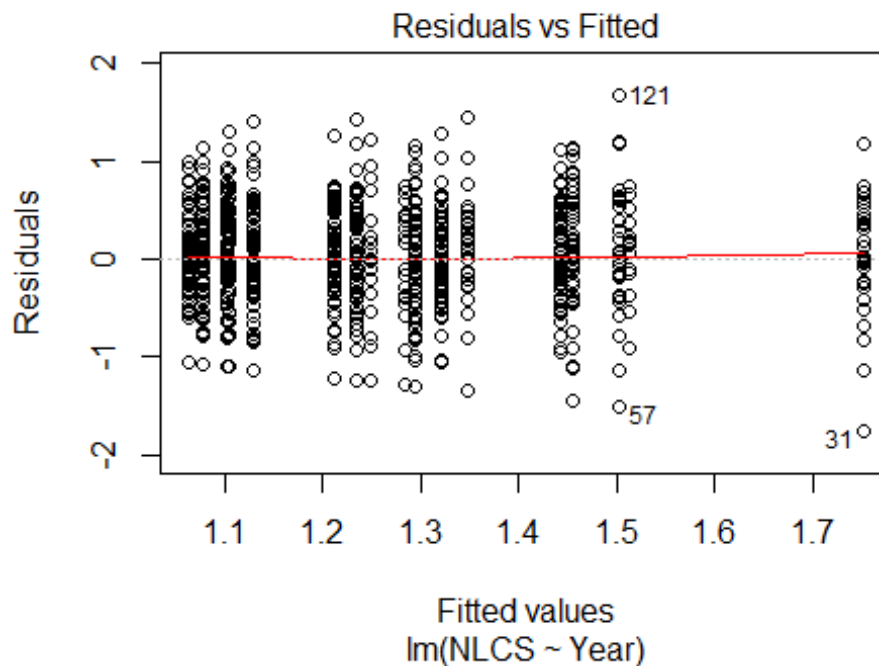
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.6195 -0.1556 0.0169 0.1683 0.7640
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0891 0.1266 8.60 4.1e-13 ***
## LastAuthorFemale1 -0.0521 0.1069 -0.49 0.63
## Year1997 -0.0981 0.1424 -0.69 0.49
## Year1998 -0.1196 0.1799 -0.66 0.51
## Year1999 -0.3056 0.2509 -1.22 0.23
## Year2000 -0.1202 0.1645 -0.73 0.47
## Year2001 0.0645 0.1629 0.40 0.69
## Year2002 -0.1429 0.1306 -1.09 0.28
## Year2003 -0.0341 0.1491 -0.23 0.82
## Year2004 0.2001 0.1328 1.51 0.14
## Year2005 -0.0591 0.1413 -0.42 0.68
## Year2007 0.4064 0.5018 0.81 0.42
```

```

## Year2008          0.0480      0.2238      0.21      0.83
## Year2009          0.0463      0.1456      0.32      0.75
## Year2010          0.0392      0.1540      0.25      0.80
## Year2011          0.1468      0.1673      0.88      0.38
## Year2012          0.2181      0.1442      1.51      0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.266
## Multiple R-squared:  0.228, Adjusted R-squared:  0.0791
## Convergence in 18 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 92 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.390  0.886  0.959  0.917  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.00e-03      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 100"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3600"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   54   70   95  108   88   84  104   82   88   96   99   94  111   21   23
## 2011 2012
##   33   27
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   34   49   63   61   46   77   65   64   72   72   78   74   20   18
## 2011 2012
##   31   26

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   29   31   48   51   59   44   69   60   57   63   63   69   66   20   18
## 2011 2012
##   30   23
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data:  NLCS by Year
## Bartlett's K-squared = 27, df = 16, p-value = 0.05
```



```
##
## Bartlett test of homogeneity of variances
##
## data:  YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2.2, df = 1, p-value = 0.1

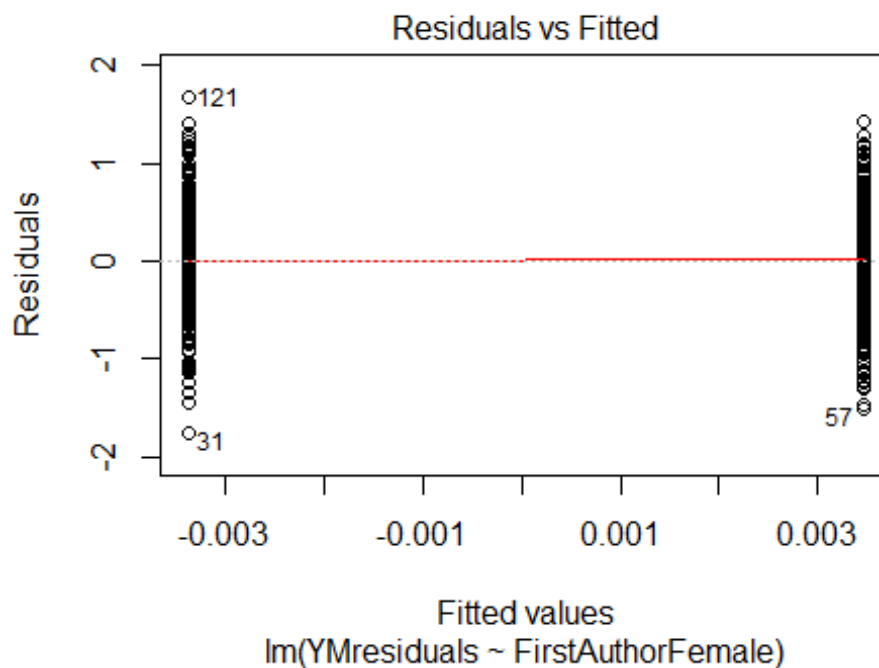
## [1] "Female first author team size 2018 geometric mean: 1.45577060189238"
## [1] "Male first author team size 2018 geometric mean: 2.68290879056234"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
```

```
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 40, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.66402887601067"
## [1] "Male last author team size 2018 geometric mean: 2.42313127540177"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

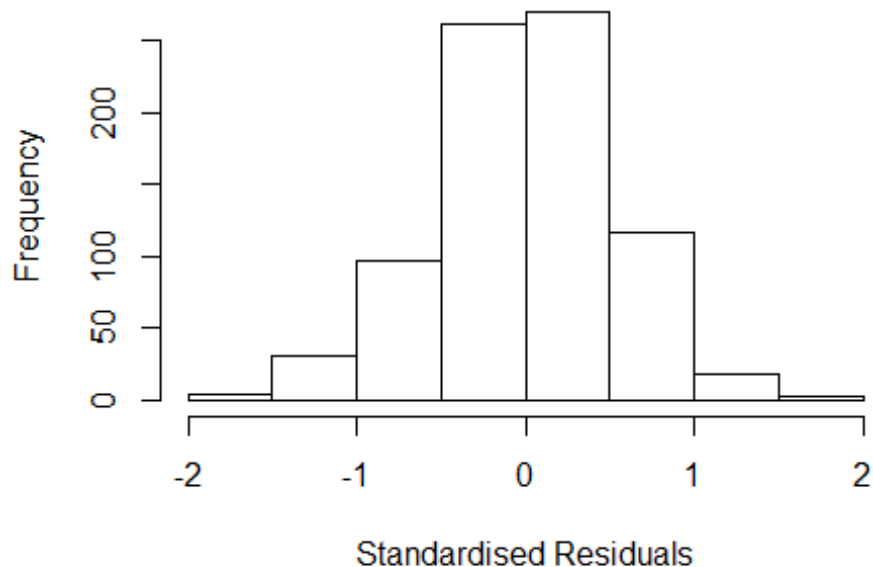


```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 44, p-value = 0.1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.307	1	1.143
LastAuthorFemale	1.432	1	1.197
UniqueAuthors	1.438	4	1.046
Year	1.592	16	1.015



## Residuals from first and last author and team size



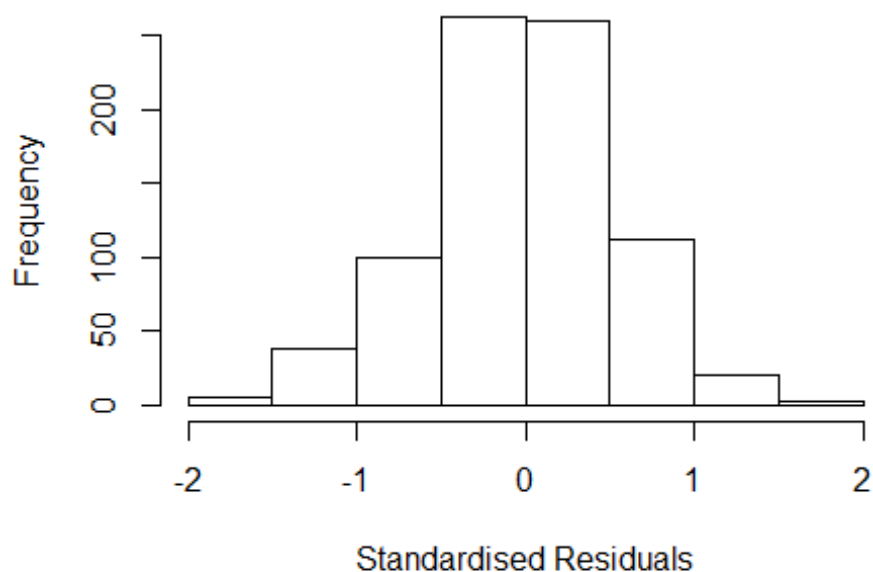
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.73201 -0.33068 0.00833 0.34787 1.60994
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7320 0.1093 15.84 < 2e-16 ***
## FirstAuthorFemale1 -0.0523 0.0427 -1.22 0.22131
## LastAuthorFemale1 0.0958 0.0444 2.16 0.03110 *
## UniqueAuthors2 0.1634 0.0504 3.24 0.00123 **
## UniqueAuthors3 0.1237 0.0558 2.22 0.02697 *
## UniqueAuthors4 0.1988 0.0680 2.92 0.00356 **
## UniqueAuthors5 0.2483 0.0776 3.20 0.00143 **
## Year1997 -0.3334 0.1720 -1.94 0.05291 .
## Year1998 -0.3138 0.1441 -2.18 0.02969 *
## Year1999 -0.4011 0.1235 -3.25 0.00121 **
```

```

## Year2000      -0.5892      0.1266      -4.66      3.8e-06 ***
## Year2001      -0.7059      0.1529      -4.62      4.5e-06 ***
## Year2002      -0.5446      0.1235      -4.41      1.2e-05 ***
## Year2003      -0.7038      0.1246      -5.65      2.3e-08 ***
## Year2004      -0.6481      0.1255      -5.16      3.1e-07 ***
## Year2005      -0.5446      0.1208      -4.51      7.5e-06 ***
## Year2006      -0.7568      0.1296      -5.84      7.7e-09 ***
## Year2007      -0.7623      0.1315      -5.80      9.8e-09 ***
## Year2008      -0.8073      0.1231      -6.56      1.0e-10 ***
## Year2009      -0.4879      0.1685      -2.90      0.00388 **
## Year2010      -0.3407      0.1390      -2.45      0.01442 *
## Year2011      -0.5022      0.1489      -3.37      0.00078 ***
## Year2012      -0.6385      0.1530      -4.17      3.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.511
## Multiple R-squared:  0.124, Adjusted R-squared:  0.0988
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 77 weights are ~= 1. The remaining 723 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.227  0.865   0.948   0.899   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.369 1      1.170
## LastAuthorFemale  1.455 1      1.206
## Year              1.176 16      1.005

```

## Residuals from first and last author



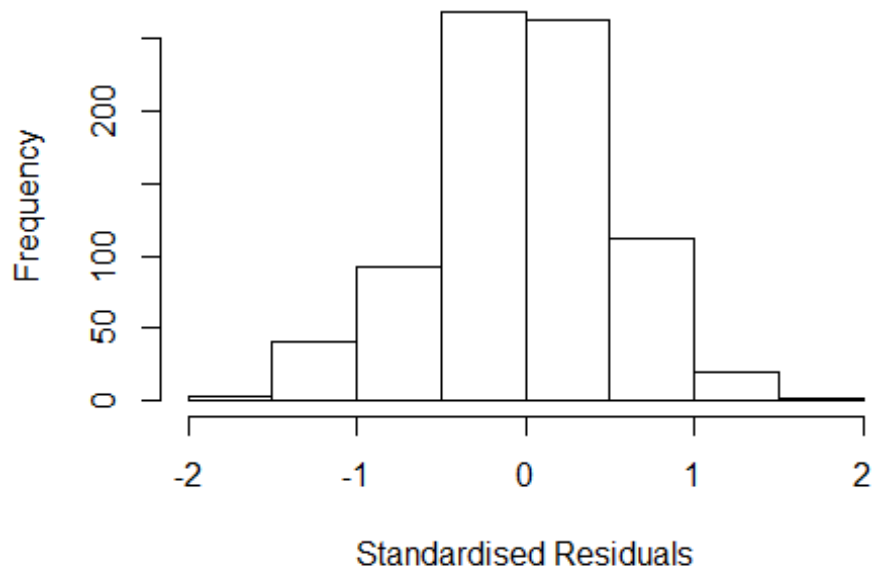
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.79325 -0.33758 -0.00266 0.35858 1.68366
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7932 0.1067 16.81 < 2e-16 ***
## FirstAuthorFemale1 -0.0499 0.0444 -1.13 0.2609
## LastAuthorFemale1 0.0766 0.0455 1.69 0.0924 .
## Year1997 -0.3049 0.1702 -1.79 0.0737 .
## Year1998 -0.2864 0.1441 -1.99 0.0473 *
## Year1999 -0.3903 0.1223 -3.19 0.0015 **
## Year2000 -0.5656 0.1240 -4.56 5.9e-06 ***
## Year2001 -0.6621 0.1478 -4.48 8.6e-06 ***
## Year2002 -0.5204 0.1224 -4.25 2.4e-05 ***
## Year2003 -0.6629 0.1219 -5.44 7.2e-08 ***
## Year2004 -0.6080 0.1249 -4.87 1.4e-06 ***
## Year2005 -0.5021 0.1174 -4.28 2.1e-05 ***
```

```

## Year2006          -0.7002      0.1278   -5.48   5.7e-08 ***
## Year2007          -0.7100      0.1263   -5.62   2.6e-08 ***
## Year2008          -0.7369      0.1205   -6.11   1.5e-09 ***
## Year2009          -0.4839      0.1651   -2.93    0.0035 **
## Year2010          -0.3065      0.1442   -2.13    0.0338 *
## Year2011          -0.4488      0.1421   -3.16    0.0016 **
## Year2012          -0.6374      0.1517   -4.20    2.9e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.524
## Multiple R-squared:  0.0996, Adjusted R-squared:  0.0788
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 62 weights are ~= 1. The remaining 738 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.218  0.874  0.954  0.904  0.988  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.076 1          1.037
## Year              1.076 16          1.002

```

## Residuals from first author



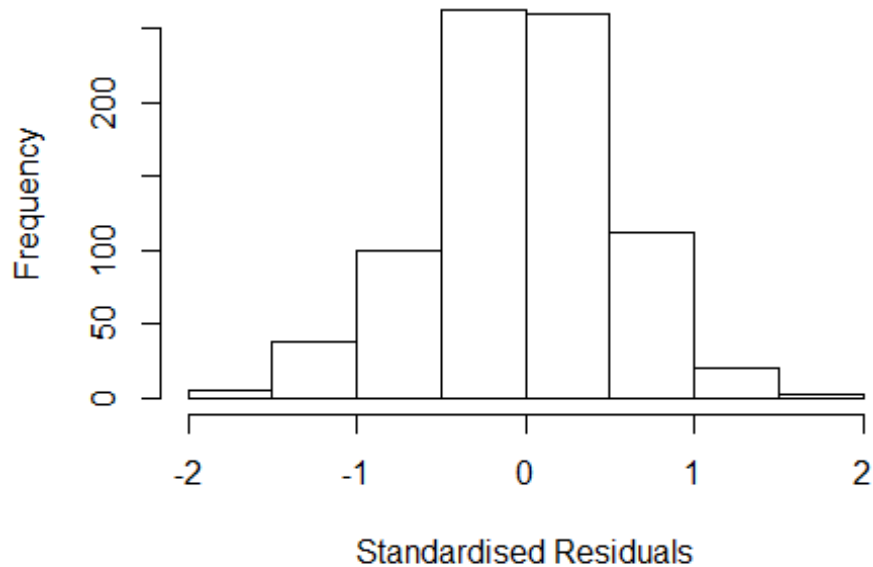
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.81158 -0.33858 -0.00736 0.36631 1.67549
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8116 0.1060 17.10 < 2e-16 ***
## FirstAuthorFemale1 -0.0106 0.0395 -0.27 0.7881
## Year1997 -0.3151 0.1676 -1.88 0.0605 .
## Year1998 -0.2872 0.1440 -1.99 0.0465 *
## Year1999 -0.3919 0.1224 -3.20 0.0014 **
## Year2000 -0.5715 0.1238 -4.62 4.5e-06 ***
## Year2001 -0.6712 0.1483 -4.53 7.0e-06 ***
## Year2002 -0.5208 0.1220 -4.27 2.2e-05 ***
## Year2003 -0.6565 0.1218 -5.39 9.2e-08 ***
## Year2004 -0.6154 0.1247 -4.94 9.7e-07 ***
## Year2005 -0.5122 0.1171 -4.37 1.4e-05 ***
## Year2006 -0.7027 0.1278 -5.50 5.2e-08 ***
```

```

## Year2007          -0.7134      0.1260    -5.66   2.1e-08 ***
## Year2008          -0.7410      0.1197    -6.19   9.6e-10 ***
## Year2009          -0.4810      0.1616    -2.98   0.0030 **
## Year2010          -0.2871      0.1460    -1.97   0.0497 *
## Year2011          -0.4457      0.1432    -3.11   0.0019 **
## Year2012          -0.6430      0.1532    -4.20   3.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.527
## Multiple R-squared:  0.0959, Adjusted R-squared:  0.0762
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 65 weights are ~= 1. The remaining 735 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.213  0.876  0.953  0.904  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.143 1          1.069
## Year            1.143 16          1.004

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.78183 -0.33265 0.00295 0.35797 1.69134
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7818 0.1065 16.74 < 2e-16 ***
## LastAuthorFemale1 0.0504 0.0404 1.25 0.2119
## Year1997 -0.3012 0.1702 -1.77 0.0773 .
## Year1998 -0.2847 0.1437 -1.98 0.0480 *
## Year1999 -0.3939 0.1232 -3.20 0.0014 **
## Year2000 -0.5646 0.1250 -4.52 7.3e-06 ***
## Year2001 -0.6678 0.1493 -4.47 8.9e-06 ***
## Year2002 -0.5214 0.1231 -4.24 2.6e-05 ***
## Year2003 -0.6597 0.1226 -5.38 9.7e-08 ***
## Year2004 -0.6108 0.1255 -4.87 1.4e-06 ***
## Year2005 -0.5049 0.1183 -4.27 2.2e-05 ***
## Year2006 -0.7021 0.1281 -5.48 5.7e-08 ***
```

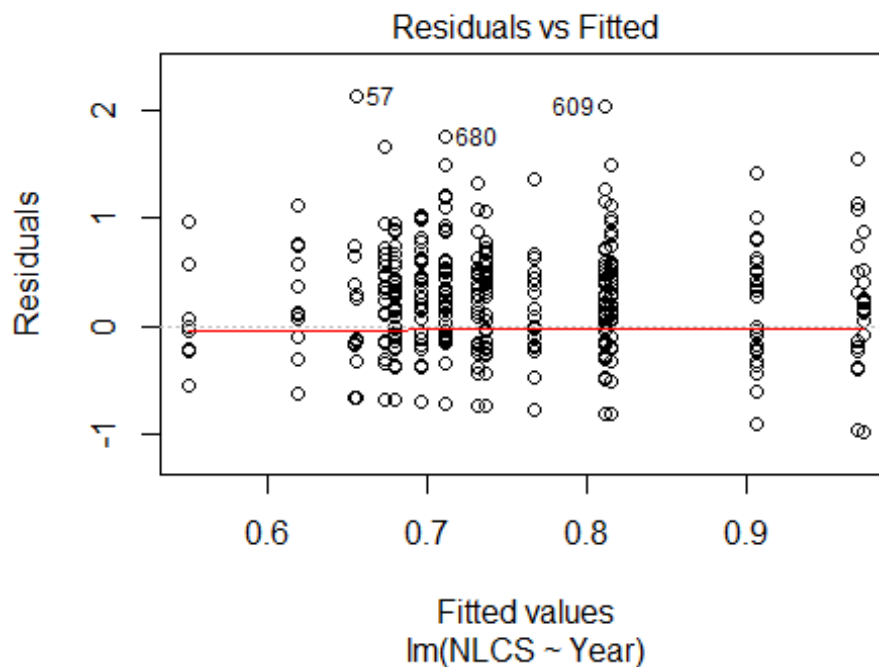
```

## Year2007          -0.7157      0.1271    -5.63  2.5e-08 ***
## Year2008          -0.7385      0.1211    -6.10  1.7e-09 ***
## Year2009          -0.4857      0.1661    -2.92   0.0036 **
## Year2010          -0.3048      0.1474    -2.07   0.0390 *
## Year2011          -0.4518      0.1434    -3.15   0.0017 **
## Year2012          -0.6420      0.1517    -4.23  2.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.525
## Multiple R-squared:  0.0982, Adjusted R-squared:  0.0786
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 69 weights are ~= 1. The remaining 731 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.225  0.871  0.952  0.903  0.987  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      1.25e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 800"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3601"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   23   18   14   27   30   46   38   26   28   60   38   42   60   57   64
## 2011 2012
##   41   79
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12    8   10   17   16   18   31   23   23   57   33   35   52   48   53
## 2011 2012

```



```
## 34 72
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 12 8 10 15 13 15 26 20 22 52 32 31 44 42 51
## 2011 2012
## 33 67
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



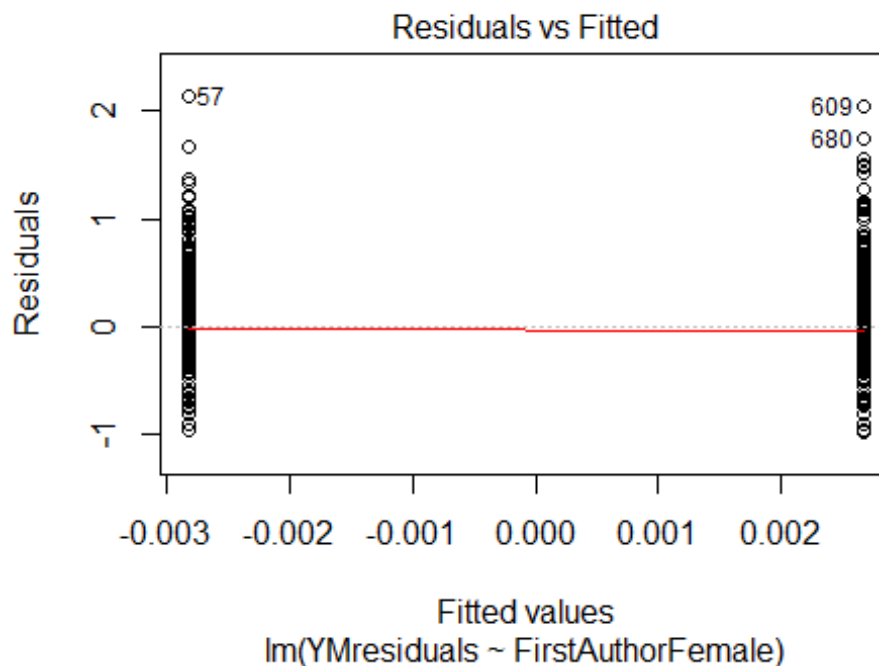
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.00041, df = 1, p-value = 1

## [1] "Female first author team size 2018 geometric mean: 1.86817216764377"
## [1] "Male first author team size 2018 geometric mean: 1.64381674866864"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

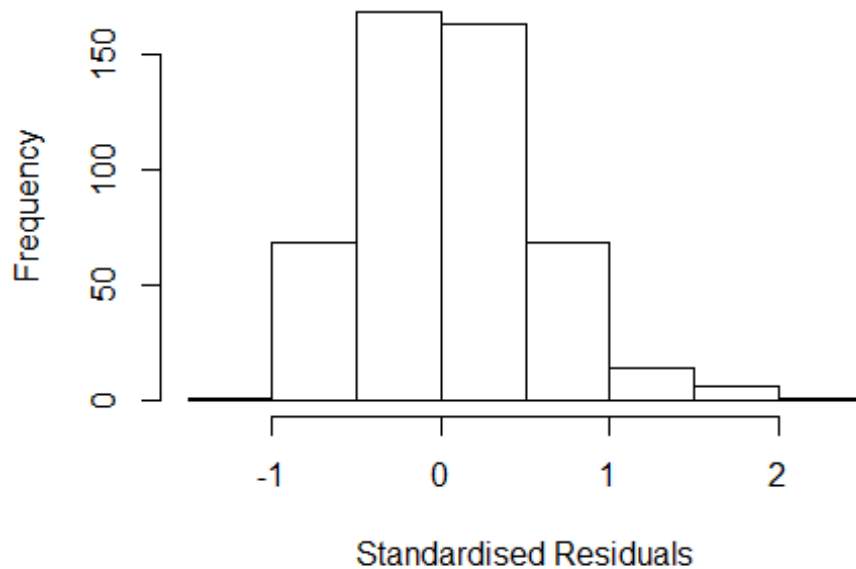
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 240, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.70936045377956"
## [1] "Male last author team size 2018 geometric mean: 1.82402813366018"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 200, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.711 1      1.308
## LastAuthorFemale  1.769 1      1.330
## UniqueAuthors    1.616 4      1.062
## Year              1.682 16     1.016
```

## Residuals from first and last author and team size



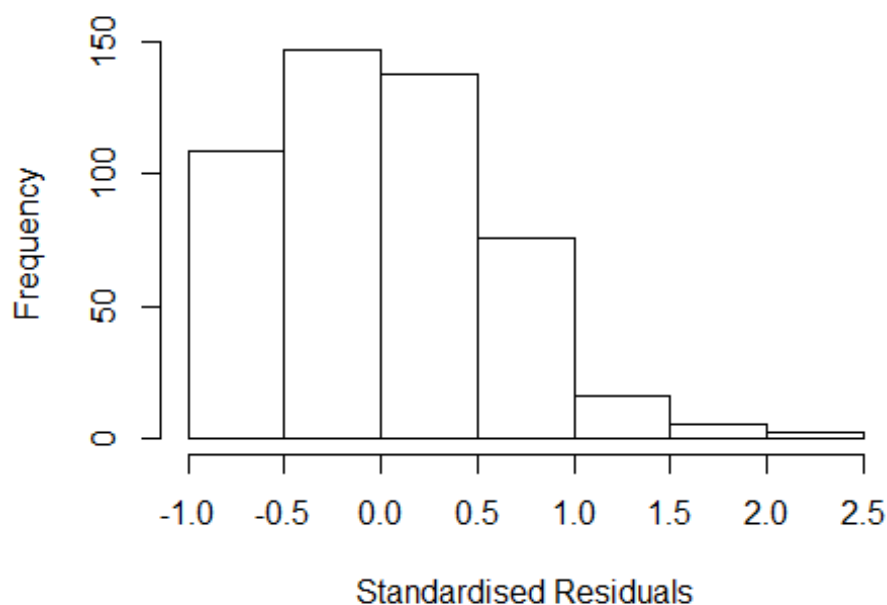
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0016 -0.4329 0.0181 0.3540 2.4296
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.448866 0.125308 3.58 0.00038 ***
## FirstAuthorFemale1 -0.036407 0.061756 -0.59 0.55578
## LastAuthorFemale1 0.010125 0.062999 0.16 0.87238
## UniqueAuthors2 0.360371 0.068814 5.24 2.5e-07 ***
## UniqueAuthors3 0.395942 0.080361 4.93 1.2e-06 ***
## UniqueAuthors4 0.382347 0.110689 3.45 0.00060 ***
## UniqueAuthors5 0.384164 0.117050 3.28 0.00111 **
## Year1997 0.016000 0.280550 0.06 0.95455
## Year1998 -0.064230 0.176310 -0.36 0.71580
## Year1999 0.192397 0.216220 0.89 0.37402
```

```

## Year2000      0.475150    0.191479    2.48  0.01343 *
## Year2001     -0.007444    0.175269   -0.04  0.96614
## Year2002      0.279070    0.183319    1.52  0.12860
## Year2003      0.064695    0.147159    0.44  0.66041
## Year2004      0.030271    0.157305    0.19  0.84748
## Year2005      0.187737    0.145033    1.29  0.19615
## Year2006      0.039274    0.152905    0.26  0.79741
## Year2007      0.013490    0.145288    0.09  0.92606
## Year2008      0.140082    0.148111    0.95  0.34474
## Year2009      0.076839    0.145900    0.53  0.59868
## Year2010      0.091231    0.142886    0.64  0.52347
## Year2011     -0.000957    0.156748   -0.01  0.99513
## Year2012      0.026250    0.144565    0.18  0.85599
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.526
## Multiple R-squared:  0.131, Adjusted R-squared:  0.09
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 46 weights are ~= 1. The remaining 447 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0008 0.8890 0.9380 0.9050 0.9830 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier          eps.x
##      1.00e-07          1.00e-07      2.03e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.715 1      1.309
## LastAuthorFemale  1.681 1      1.297
## Year              1.119 16      1.004

```

## Residuals from first and last author

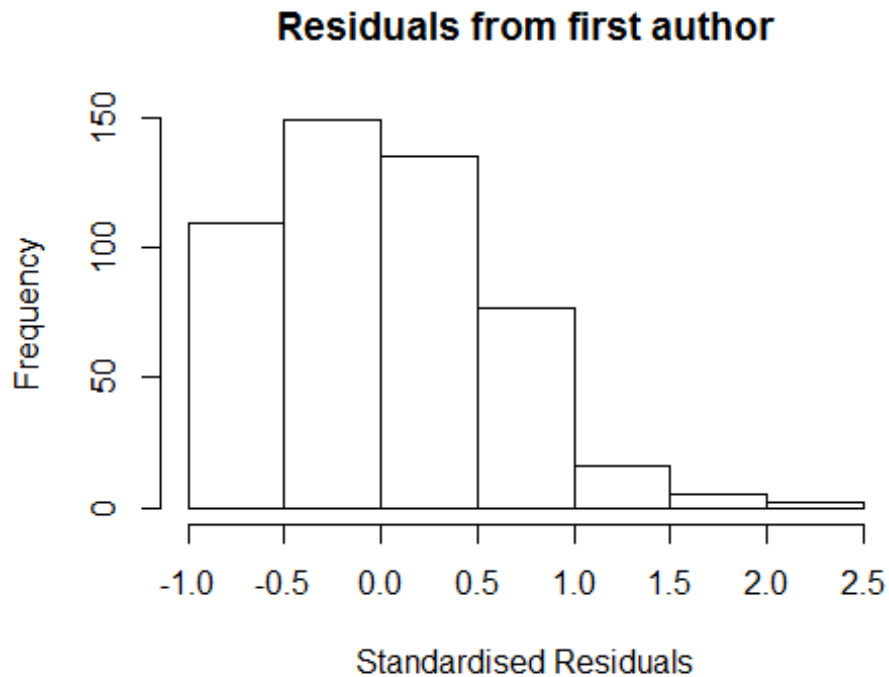


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9567 -0.4001 -0.0528 0.4159 2.3445
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.5344 0.1444 3.70 0.00024 ***
## FirstAuthorFemale1 -0.0164 0.0658 -0.25 0.80359
## LastAuthorFemale1 0.0138 0.0661 0.21 0.83491
## Year1997 0.1117 0.2443 0.46 0.64783
## Year1998 -0.0883 0.1960 -0.45 0.65265
## Year1999 0.2079 0.2126 0.98 0.32865
## Year2000 0.4224 0.2181 1.94 0.05340 .
## Year2001 0.0416 0.1899 0.22 0.82656
## Year2002 0.3510 0.1885 1.86 0.06314 .
## Year2003 0.1401 0.1758 0.80 0.42598
## Year2004 0.0315 0.1697 0.19 0.85300
## Year2005 0.2405 0.1642 1.46 0.14384
```

```

## Year2006          0.1645      0.1793      0.92  0.35957
## Year2007          0.1159      0.1629      0.71  0.47718
## Year2008          0.1707      0.1630      1.05  0.29550
## Year2009          0.1015      0.1656      0.61  0.54035
## Year2010          0.2085      0.1597      1.31  0.19222
## Year2011          0.1346      0.1733      0.78  0.43780
## Year2012          0.1553      0.1623      0.96  0.33904
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.602
## Multiple R-squared:  0.0254, Adjusted R-squared:  -0.0116
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 466 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0952 0.8850 0.9540 0.9210 0.9890 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.03e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.076 1      1.037
## Year              1.076 16      1.002

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -0.9580 -0.4024 -0.0518 0.4131 2.3464
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.53842 0.14350 3.75 0.0002 ***
## FirstAuthorFemale1 -0.00845 0.05212 -0.16 0.8713
## Year1997 0.10979 0.24402 0.45 0.6530
## Year1998 -0.08834 0.19657 -0.45 0.6533
## Year1999 0.20804 0.21311 0.98 0.3295
## Year2000 0.41957 0.21725 1.93 0.0540 .
## Year2001 0.03954 0.18963 0.21 0.8349
## Year2002 0.35079 0.18823 1.86 0.0630 .
## Year2003 0.13946 0.17558 0.79 0.4274
## Year2004 0.02917 0.16923 0.17 0.8632
## Year2005 0.23876 0.16383 1.46 0.1457
## Year2006 0.16300 0.17892 0.91 0.3627
```

```

## Year2007          0.11523    0.16274    0.71    0.4793
## Year2008          0.16894    0.16291    1.04    0.3003
## Year2009          0.10124    0.16556    0.61    0.5412
## Year2010          0.20823    0.15956    1.31    0.1925
## Year2011          0.13421    0.17306    0.78    0.4384
## Year2012          0.15338    0.16150    0.95    0.3427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.605
## Multiple R-squared:  0.0252, Adjusted R-squared:  -0.00965
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 27 weights are ~= 1. The remaining 466 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0995 0.8870 0.9540 0.9220 0.9900 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.03e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.054 1      1.027
## Year      1.054 16      1.002
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -0.9563 -0.3989 -0.0511  0.4152  2.3373

```



```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.53018    0.14344   3.70 0.00024 ***
## LastAuthorFemale1 0.00449    0.05244   0.09 0.93177
## Year1997        0.11743    0.24238   0.48 0.62826
## Year1998       -0.08399    0.19633  -0.43 0.66899
## Year1999        0.21229    0.21189   1.00 0.31692
## Year2000        0.42162    0.21757   1.94 0.05323 .
## Year2001        0.04350    0.18960   0.23 0.81862
## Year2002        0.35169    0.18838   1.87 0.06253 .
## Year2003        0.14154    0.17583   0.80 0.42125
## Year2004        0.03033    0.16945   0.18 0.85801
## Year2005        0.24094    0.16430   1.47 0.14318
## Year2006        0.16474    0.17870   0.92 0.35707
## Year2007        0.11690    0.16259   0.72 0.47249
## Year2008        0.17109    0.16345   1.05 0.29577
## Year2009        0.10330    0.16592   0.62 0.53386
## Year2010        0.21094    0.15986   1.32 0.18762
## Year2011        0.13516    0.17310   0.78 0.43530
## Year2012        0.15591    0.16221   0.96 0.33694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.609
## Multiple R-squared:  0.0251, Adjusted R-squared:  -0.00978
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 31 weights are ~= 1. The remaining 462 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.107 0.888  0.954  0.922  0.989  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.03e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 493"
## [1] ""

```

```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3602"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    4   10    5    5    7   16   12    8   19   19   31   26   20   30
## 2012
##    15
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    0    5    5    5    6   14    9    5   14   17   25   22   16   26
## 2012
##     8
##
## 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
##    1    0    5    5    4    5   14    9    5   14   17   20   19   15   26
## 2012
##     7
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.95611498831023"
## [1] "Male first author team size 2018 geometric mean: 2.90072293183791"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

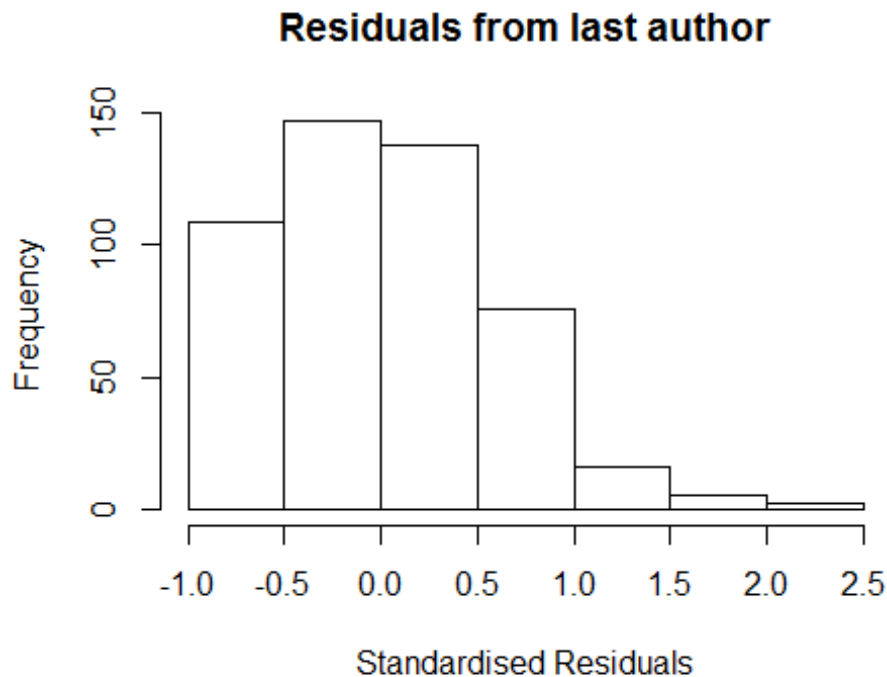
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.71140042032231"
## [1] "Male last author team size 2018 geometric mean: 3.12031850602576"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 33, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

```

```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale  7.986  1      2.826
## LastAuthorFemale  7.846  1      2.801
## UniqueAuthors    11.531  4      1.357
## Year              58.520 14      1.156

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2971 -0.3183  0.0253  0.3424  1.6357
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.63000    0.00000      Inf <2e-16 ***
```

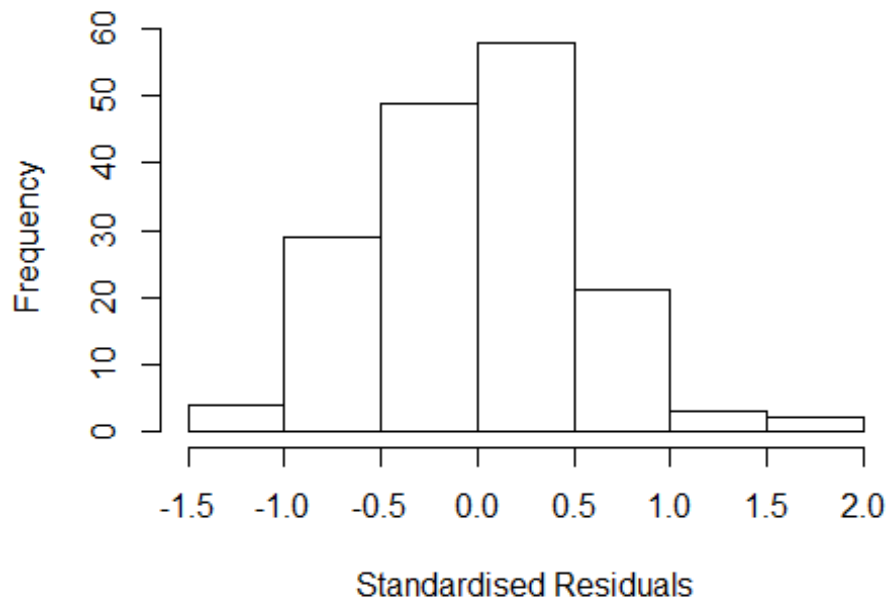
```

## FirstAuthorFemale1  0.11027    0.12096    0.91    0.3635
## LastAuthorFemale1  -0.19442    0.12785   -1.52    0.1305
## UniqueAuthors2     0.28207    0.13847    2.04    0.0435 *
## UniqueAuthors3     0.32272    0.18010    1.79    0.0752 .
## UniqueAuthors4     0.21798    0.18185    1.20    0.2326
## UniqueAuthors5     0.25558    0.15312    1.67    0.0973 .
## Year1999            -0.36734    0.11386   -3.23    0.0016 **
## Year2000            -0.25188    0.26540   -0.95    0.3442
## Year2001            -0.07371    0.38808   -0.19    0.8496
## Year2002            -0.00734    0.35348   -0.02    0.9835
## Year2003             0.00646    0.17430    0.04    0.9705
## Year2004             0.09689    0.16185    0.60    0.5503
## Year2005             0.48910    0.18023    2.71    0.0075 **
## Year2006             0.54826    0.22919    2.39    0.0180 *
## Year2007             0.34438    0.15511    2.22    0.0280 *
## Year2008             0.34003    0.19456    1.75    0.0826 .
## Year2009             0.08283    0.18168    0.46    0.6491
## Year2010            -0.02482    0.18807   -0.13    0.8952
## Year2011             0.02330    0.18836    0.12    0.9017
## Year2012             0.15536    0.35344    0.44    0.6609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.265, Adjusted R-squared:  0.164
## Convergence in 15 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 150 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.325  0.874  0.956  0.908  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample    max.it    best.r.s    k.fast.s    k.max maxit.scale
##      500         50         2         1        1000         200
##   trace.lev    mts  compute.rd
##      0        1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"

```

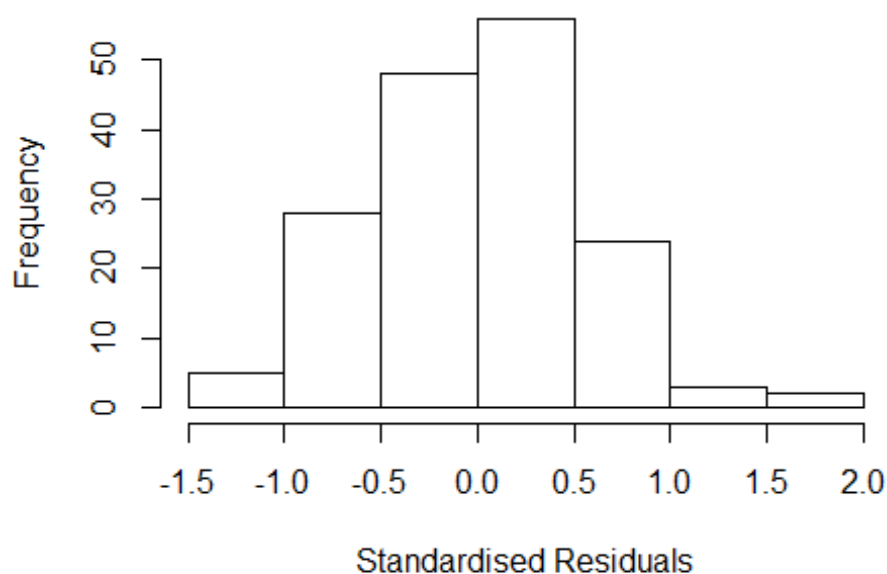
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead
```

## Residuals from first and last author and team size



##		GVIF	Df	GVIF^(1/(2*Df))
##	FirstAuthorFemale	6.841	1	2.615
##	LastAuthorFemale	6.007	1	2.451
##	Year	9.669	14	1.084

## Residuals from first and last author



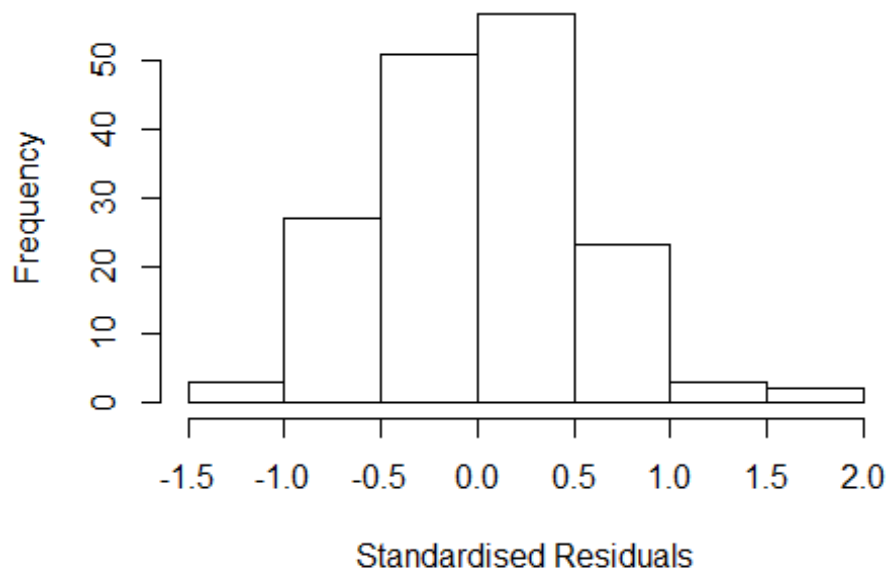
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14443 -0.40603 0.00576 0.38162 1.62397
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6300 0.0000 Inf < 2e-16 ***
## FirstAuthorFemale1 0.1224 0.1168 1.05 0.29650
## LastAuthorFemale1 -0.2311 0.1165 -1.98 0.04915 *
## Year1999 -0.3525 0.1143 -3.08 0.00243 **
## Year2000 -0.2186 0.2661 -0.82 0.41264
## Year2001 0.0232 0.3116 0.07 0.94078
## Year2002 0.1697 0.3349 0.51 0.61316
## Year2003 0.0428 0.1648 0.26 0.79550
## Year2004 0.1460 0.1712 0.85 0.39500
## Year2005 0.6635 0.2040 3.25 0.00142 **
## Year2006 0.7596 0.1839 4.13 6e-05 ***
## Year2007 0.4909 0.1450 3.39 0.00091 ***
```

```

## Year2008          0.6075      0.1679      3.62  0.00041 ***
## Year2009          0.3331      0.1395      2.39  0.01821 *
## Year2010          0.1802      0.1561      1.15  0.25019
## Year2011          0.2365      0.1531      1.55  0.12439
## Year2012          0.3921      0.3536      1.11  0.26935
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.571
## Multiple R-squared:  0.225, Adjusted R-squared:  0.142
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 146 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.398  0.882  0.946  0.914  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 5.229 1          2.287
## Year              5.229 14          1.061

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.1748 -0.3817 0.0315 0.3612 1.7097
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.30e-01 1.08e-07 5.84e+06 < 2e-16 ***
## FirstAuthorFemale1 1.78e-02 1.04e-01 1.70e-01 0.86421
## Year1999 -4.29e-01 1.07e-01 -4.02e+00 9.1e-05 ***
## Year2000 -3.14e-01 2.70e-01 -1.16e+00 0.24712
## Year2001 -9.28e-02 3.50e-01 -2.70e-01 0.79110
## Year2002 5.79e-02 2.91e-01 2.00e-01 0.84257
## Year2003 -3.26e-02 1.55e-01 -2.10e-01 0.83332
## Year2004 6.68e-02 1.81e-01 3.70e-01 0.71237
## Year2005 6.33e-01 1.80e-01 3.51e+00 0.00059 ***
## Year2006 6.73e-01 1.86e-01 3.62e+00 0.00040 ***
## Year2007 4.94e-01 1.48e-01 3.33e+00 0.00109 **
## Year2008 5.45e-01 1.64e-01 3.33e+00 0.00110 **
```

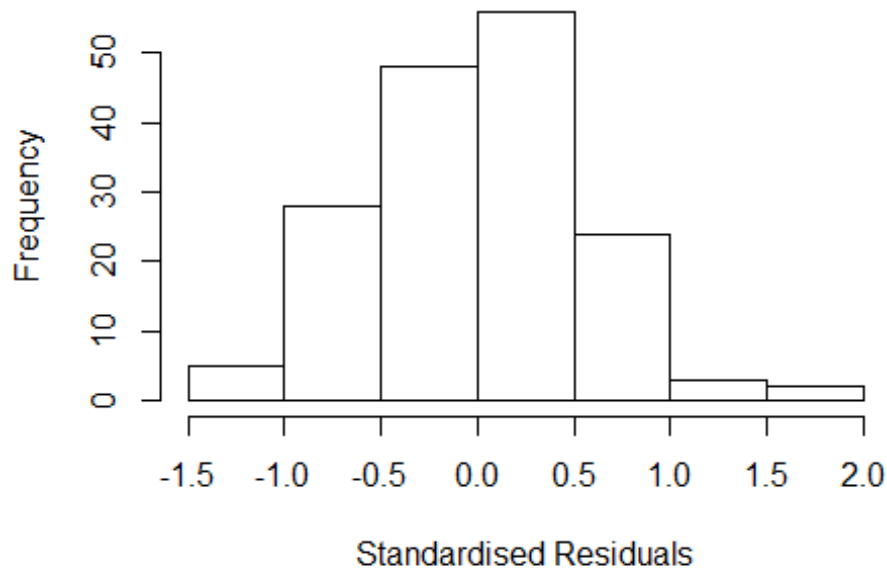


```

## Year2009          2.74e-01   1.39e-01   1.97e+00   0.05053 .
## Year2010          1.15e-01   1.45e-01   8.00e-01   0.42615
## Year2011          1.78e-01   1.58e-01   1.13e+00   0.26014
## Year2012          2.45e-01   2.73e-01   9.00e-01   0.37081
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.202, Adjusted R-squared:  0.122
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 148 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.349  0.877  0.953  0.912  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 4.626 1          2.151
## Year             4.626 14          1.056

```

## Residuals from last author



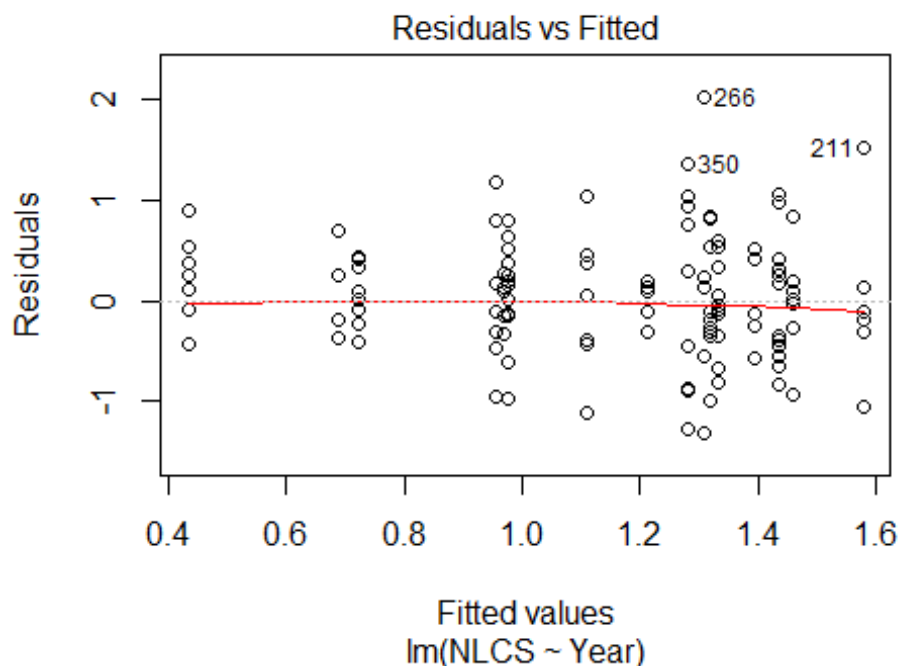
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.18360 -0.39134 0.00794 0.38556 1.65259
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.30e-01 9.02e-08 6.99e+06 < 2e-16 ***
## LastAuthorFemale1 -1.80e-01 1.02e-01 -1.76e+00 0.08098 .
## Year1999 -3.09e-01 1.10e-01 -2.80e+00 0.00572 **
## Year2000 -1.66e-01 2.52e-01 -6.60e-01 0.51012
## Year2001 3.87e-02 3.25e-01 1.20e-01 0.90518
## Year2002 2.20e-01 3.04e-01 7.20e-01 0.47121
## Year2003 8.54e-02 1.67e-01 5.10e-01 0.61029
## Year2004 1.90e-01 1.64e-01 1.16e+00 0.24818
## Year2005 7.15e-01 1.94e-01 3.69e+00 0.00032 ***
## Year2006 8.20e-01 1.72e-01 4.75e+00 4.6e-06 ***
## Year2007 5.54e-01 1.27e-01 4.37e+00 2.3e-05 ***
## Year2008 6.67e-01 1.52e-01 4.38e+00 2.2e-05 ***
```

```

## Year2009          3.92e-01  1.28e-01  3.05e+00  0.00268 **
## Year2010          2.46e-01  1.44e-01  1.71e+00  0.08978 .
## Year2011          2.99e-01  1.46e-01  2.05e+00  0.04215 *
## Year2012          4.11e-01  3.47e-01  1.19e+00  0.23740
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.57
## Multiple R-squared:  0.22,   Adjusted R-squared:  0.142
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 146 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.381  0.875  0.946  0.913  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      6.02e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 166"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3603"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   44   41   29   29   27   21   18   13    9    7    9   12    9   11    7
## 2011 2012
##   16   11
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   15   12    8    5    6    5    6   11    6    5    6    9    5    8    7
## 2011 2012
##   14   10
##

```

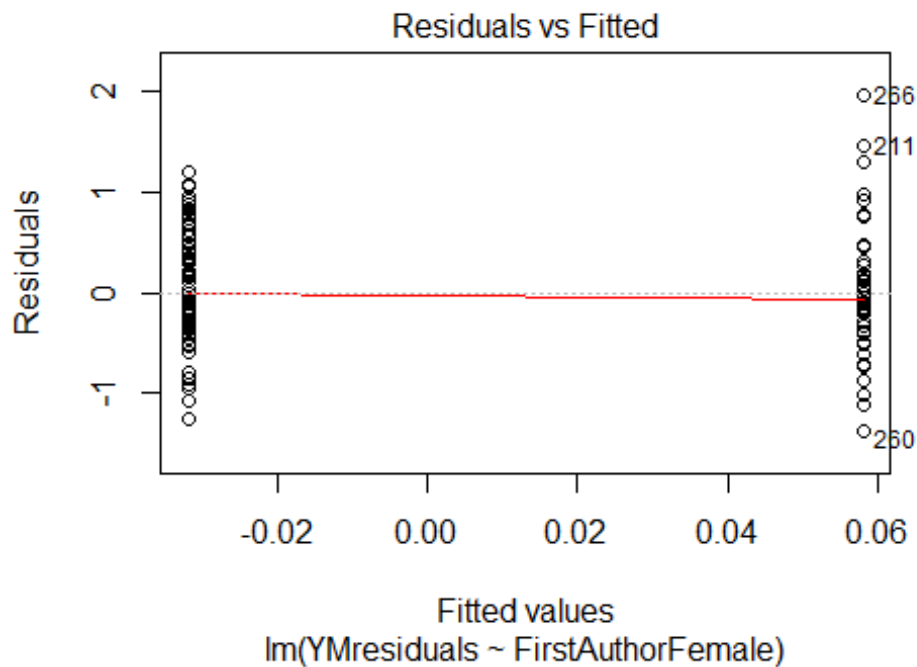
```
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 12 6 4 6 5 6 4 6 4 5 7 4 8 6
## 2011 2012
## 14 10
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 36, df = 16, p-value = 0.003
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.7, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 3"
## [1] "Male first author team size 2018 geometric mean: 3.36042145371267"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
##
## Wilcoxon rank sum test with continuity correction
##
```

```
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.48140474655716"
## [1] "Male last author team size 2018 geometric mean: 2.70192007704123"

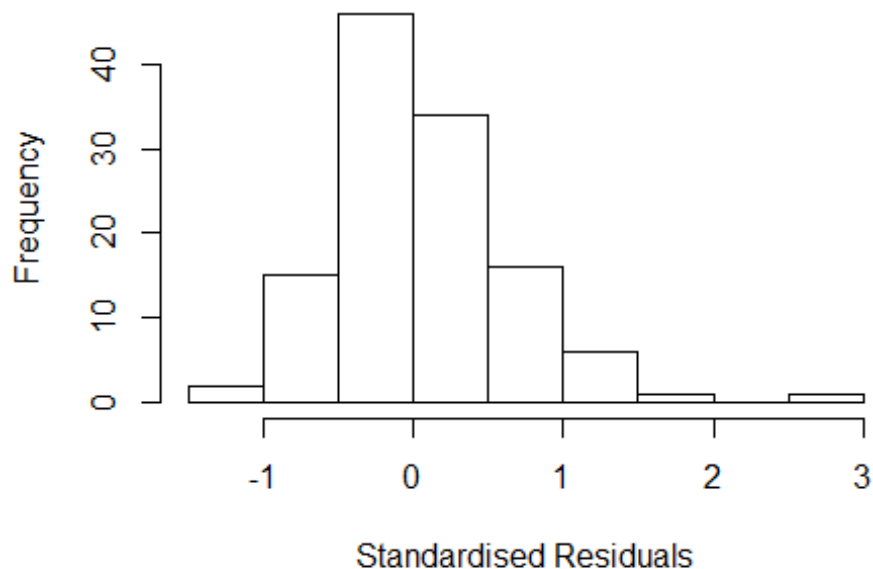
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 12, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	4.028	1	2.007
LastAuthorFemale	4.028	1	2.007
UniqueAuthors	18.281	4	1.438
Year	44.837	16	1.126

## Residuals from first and last author and team size



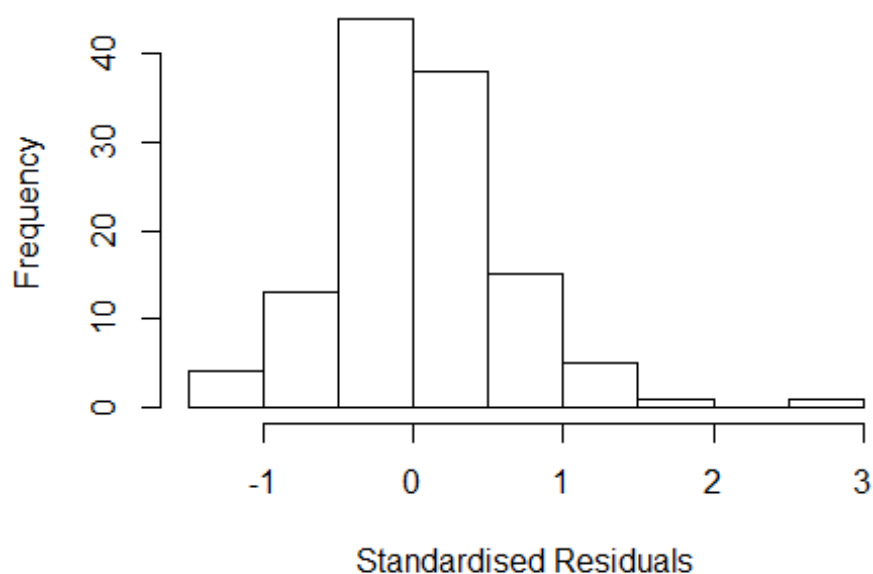
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 266 33847319171 3.333 2006      1210      2      2.614
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1719 -0.3141 -0.0122  0.2909  2.6144
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.31409    0.11917   2.64 0.00976 **
## FirstAuthorFemale1 0.00739    0.16059   0.05 0.96338
## LastAuthorFemale1 -0.04326    0.16840  -0.26 0.79781
## UniqueAuthors2    0.24173    0.17254   1.40 0.16439
## UniqueAuthors3    0.54075    0.19170   2.82 0.00580 **
## UniqueAuthors4    0.50261    0.31245   1.61 0.11092
## UniqueAuthors5    0.45013    0.17377   2.59 0.01105 *
## Year1997          0.24634    0.12294   2.00 0.04786 *
## Year1998          0.68188    0.41466   1.64 0.10329
## Year1999          0.47015    0.25707   1.83 0.07046 .
```

```

## Year2000      1.10155      0.30799      3.58  0.00054 ***
## Year2001      0.70614      0.13261      5.33  6.4e-07 ***
## Year2002      0.87430      0.35794      2.44  0.01637 *
## Year2003      0.75540      0.21836      3.46  0.00080 ***
## Year2004      0.74308      0.32773      2.27  0.02557 *
## Year2005      0.60090      0.15581      3.86  0.00021 ***
## Year2006      0.15537      0.23192      0.67  0.50448
## Year2007      0.91821      0.33854      2.71  0.00789 **
## Year2008      0.53522      0.31351      1.71  0.09095 .
## Year2009      0.82200      0.25302      3.25  0.00159 **
## Year2010      0.59385      0.35139      1.69  0.09421 .
## Year2011      0.39851      0.20427      1.95  0.05392 .
## Year2012      0.57178      0.37308      1.53  0.12859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.537
## Multiple R-squared:  0.336, Adjusted R-squared:  0.187
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## observation 72 is an outlier with |weight| = 0 ( < 0.00083);
## 12 weights are ~= 1. The remaining 108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.285  0.866  0.963   0.906  0.990   0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           8.26e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample max.it best.r.s k.fast.s k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev mts compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 3.497 1 1.870
## LastAuthorFemale 4.106 1 2.026
## Year 5.796 16 1.056

```

## Residuals from first and last author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 266 33847319171 3.333 2006    1210      2    2.541
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## --> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2378 -0.3465 -0.0131  0.3327  2.5406
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3877    0.1101   3.52 0.00064 ***
## FirstAuthorFemale1  0.0375    0.1659   0.23 0.82172
## LastAuthorFemale1 -0.0350    0.1804  -0.19 0.84669
## Year1997         0.3276    0.1322   2.48 0.01488 *
## Year1998         0.5952    0.4174   1.43 0.15694
## Year1999         0.3856    0.2524   1.53 0.12970
## Year2000         1.0268    0.2984   3.44 0.00084 ***
## Year2001         0.8171    0.1364   5.99 3.2e-08 ***
## Year2002         1.0262    0.3082   3.33 0.00121 **
## Year2003         0.7582    0.3041   2.49 0.01427 *
## Year2004         0.9929    0.3196   3.11 0.00245 **
## Year2005         0.5590    0.1976   2.83 0.00562 **
```

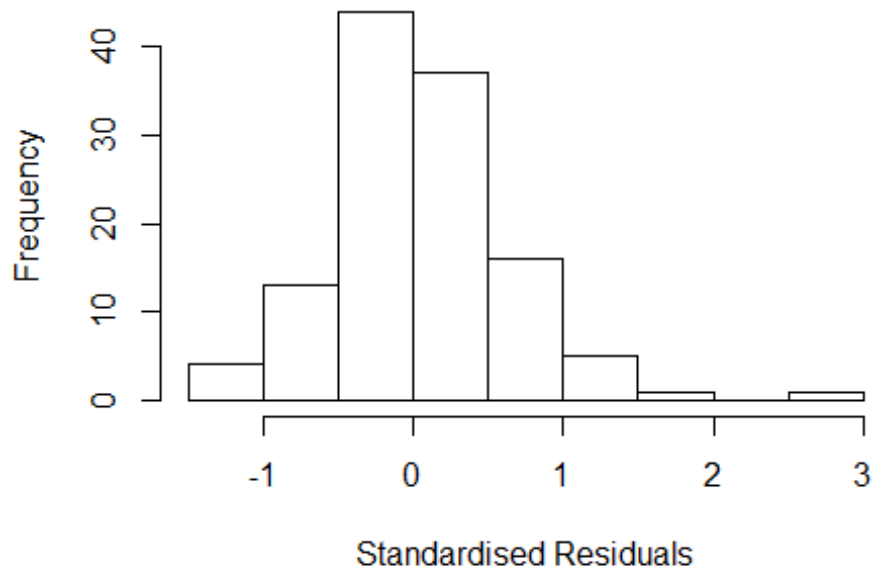


```

## Year2006          0.3672      0.3175      1.16  0.25018
## Year2007          1.0022      0.3235      3.10  0.00251 **
## Year2008          0.8944      0.2491      3.59  0.00051 ***
## Year2009          1.0740      0.1951      5.51  2.8e-07 ***
## Year2010          0.6224      0.3478      1.79  0.07652 .
## Year2011          0.6179      0.1962      3.15  0.00215 **
## Year2012          0.8501      0.6163      1.38  0.17080
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.54
## Multiple R-squared:  0.269, Adjusted R-squared:  0.14
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## observation 72 is an outlier with |weight| = 0 ( < 0.00083);
## 12 weights are ~= 1. The remaining 108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.304  0.826   0.954   0.898   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.26e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 2.495 1          1.580
## Year              2.495 16          1.029

```

## Residuals from first author



```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 266 33847319171 3.333 2006      1210      2      2.541
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min        1Q    Median        3Q        Max
## -1.23530 -0.35487 -0.00648  0.33587  2.55169
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.3861     0.1116   3.46  0.00079 ***
## FirstAuthorFemale1 0.0261     0.1427   0.18  0.85509
## Year1997          0.3262     0.1327   2.46  0.01562 *
## Year1998          0.5936     0.4050   1.47  0.14573
## Year1999          0.3812     0.2475   1.54  0.12649
## Year2000          1.0290     0.2906   3.54  0.00060 ***
## Year2001          0.8161     0.1392   5.86  5.5e-08 ***
## Year2002          1.0457     0.3116   3.36  0.00111 **
## Year2003          0.7515     0.3017   2.49  0.01436 *
## Year2004          0.9957     0.3088   3.22  0.00169 **
## Year2005          0.5427     0.1806   3.00  0.00334 **
## Year2006          0.3691     0.3391   1.09  0.27891
```

```

## Year2007          0.9886      0.3089      3.20  0.00182 **
## Year2008          0.8715      0.2129      4.09  8.4e-05 ***
## Year2009          1.0697      0.1959      5.46  3.3e-07 ***
## Year2010          0.6157      0.3313      1.86  0.06597 .
## Year2011          0.6059      0.1893      3.20  0.00182 **
## Year2012          0.8492      0.5681      1.49  0.13804
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.566
## Multiple R-squared:  0.265, Adjusted R-squared:  0.144
## Convergence in 22 IRWLS iterations
##
## Robustness weights:
## 12 weights are ~= 1. The remaining 109 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0052 0.8400 0.9580 0.8980 0.9870 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.26e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 3.065 1      1.751
## Year      3.065 16      1.036
##
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId NLCS Year OneField Fields residuals
## 266 33847319171 3.333 2006      1210      2      2.541
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.23766 -0.34128 -0.00563  0.33434  2.56037

```

```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.3931    0.1038   3.79 0.00026 ***
## LastAuthorFemale1 -0.0131    0.1562  -0.08 0.93349
## Year1997        0.3281    0.1321   2.48 0.01459 *
## Year1998        0.5962    0.4142   1.44 0.15303
## Year1999        0.3848    0.2550   1.51 0.13434
## Year2000        1.0224    0.2952   3.46 0.00078 ***
## Year2001        0.8226    0.1368   6.01 2.8e-08 ***
## Year2002        1.0218    0.2963   3.45 0.00082 ***
## Year2003        0.7533    0.2952   2.55 0.01218 *
## Year2004        0.9963    0.3051   3.27 0.00149 **
## Year2005        0.5655    0.1947   2.90 0.00450 **
## Year2006        0.3795    0.3153   1.20 0.23138
## Year2007        1.0060    0.3243   3.10 0.00248 **
## Year2008        0.8734    0.2276   3.84 0.00021 ***
## Year2009        1.0787    0.1938   5.57 2.1e-07 ***
## Year2010        0.6247    0.3494   1.79 0.07671 .
## Year2011        0.6163    0.1968   3.13 0.00227 **
## Year2012        0.8446    0.6191   1.36 0.17545
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.544
## Multiple R-squared:  0.268, Adjusted R-squared:  0.147
## Convergence in 25 IRWLS iterations
##
## Robustness weights:
## observation 72 is an outlier with |weight| = 0 ( < 0.00083);
## 12 weights are ~= 1. The remaining 108 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.302  0.828  0.955  0.899  0.987  0.999
## Algorithmic parameters:
##           tuning.chi          bb          tuning.psi          refine.tol
##           1.55e+00          5.00e-01          4.69e+00          1.00e-07
##           rel.tol          solve.tol          eps.outlier          eps.x
##           1.00e-07          1.00e-07          8.26e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi          subsampling          cov
##           "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 121"

```

```

## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3604"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1998 1999 2000 2001 2005 2006 2007 2008 2009 2011 2012
##    1    3    2    2    1    5    2    2    3    4    2
##
## 1998 1999 2000 2001 2005 2006 2007 2008 2009 2011 2012
##    1    3    2    1    1    4    1    1    2    1    0
##
## 1998 1999 2000 2001 2005 2006 2007 2008 2009 2011 2012
##    1    3    2    1    1    4    1    1    2    1    0
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.86193816221051"
## [1] "Male first author team size 2018 geometric mean: 10"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 0, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2"
## [1] "Male last author team size 2018 geometric mean: 4.68069463864143"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 17"

```

```

## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3605"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 14 8 5 10 14 21 27 28 13 45 54 62 21 25 17
## 2011 2012
## 23 26
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 2 2 3 1 4 3 14 16 12 39 44 37 14 18 14
## 2011 2012
## 21 16
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 1 2 2 1 4 3 14 14 8 36 39 34 14 14 12
## 2011 2012
## 18 14
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 2.54365986748923"
## [1] "Male first author team size 2018 geometric mean: 2.62647461529191"

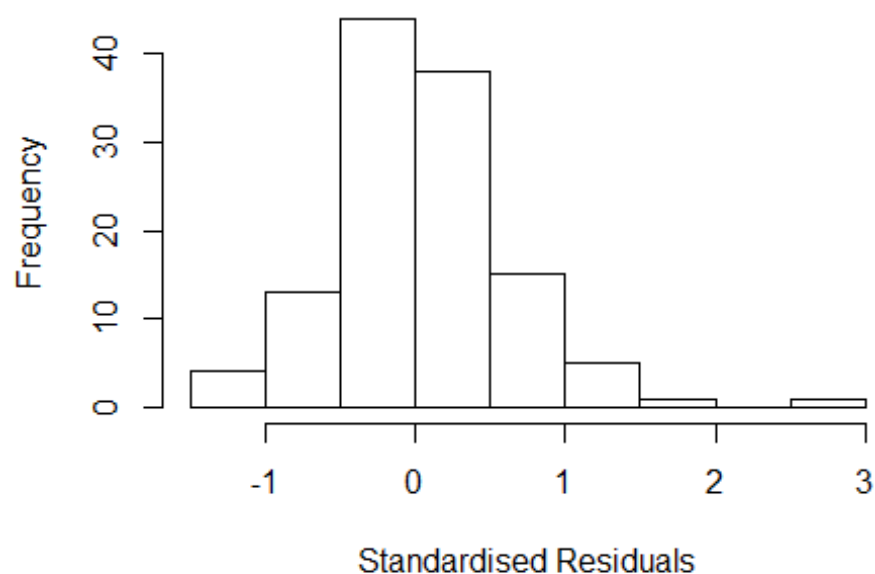
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.7
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.72096526527247"
## [1] "Male last author team size 2018 geometric mean: 2.53598179475474"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

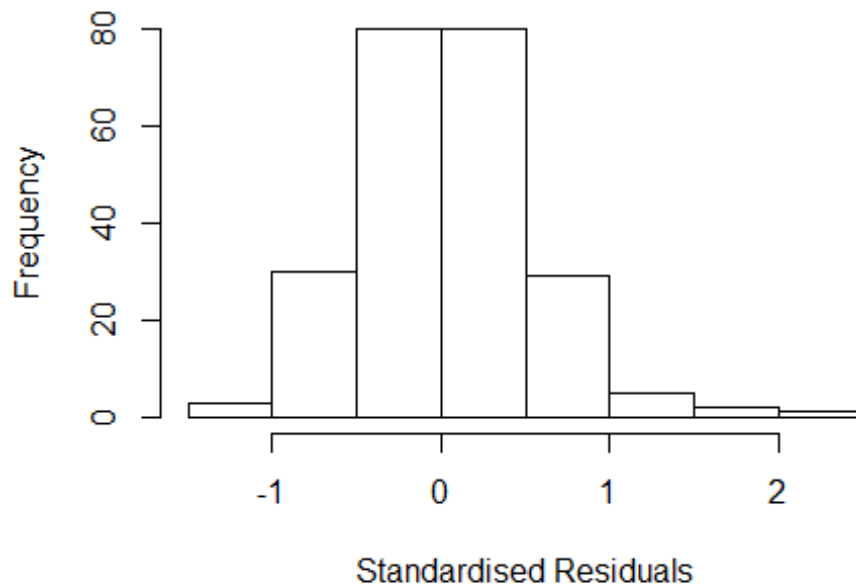
```

## Residuals from last author



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 210, p-value = 0.9
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 8.045e+12  1      2.836e+06
## LastAuthorFemale  3.926e+00  1      1.981e+00
## UniqueAuthors    1.057e+15  4      7.551e+01
## Year              5.120e+15 16      3.097e+00
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.057 -0.308 0.012 0.280 2.065
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.5486 0.1048 -5.23 4.1e-07 ***
## FirstAuthorFemale1 -0.0804 0.0723 -1.11 0.26733
## LastAuthorFemale1 0.2064 0.0768 2.69 0.00774 **
## UniqueAuthors2 0.2753 0.0887 3.10 0.00219 **
## UniqueAuthors3 0.4385 0.1157 3.79 0.00020 ***
## UniqueAuthors4 0.5486 0.1048 5.23 4.1e-07 ***
## UniqueAuthors5 0.3779 0.0994 3.80 0.00019 ***
## Year1997 1.6641 0.1066 15.61 < 2e-16 ***
## Year1998 1.6234 0.3606 4.50 1.1e-05 ***
## Year1999 1.4847 0.1311 11.33 < 2e-16 ***
```

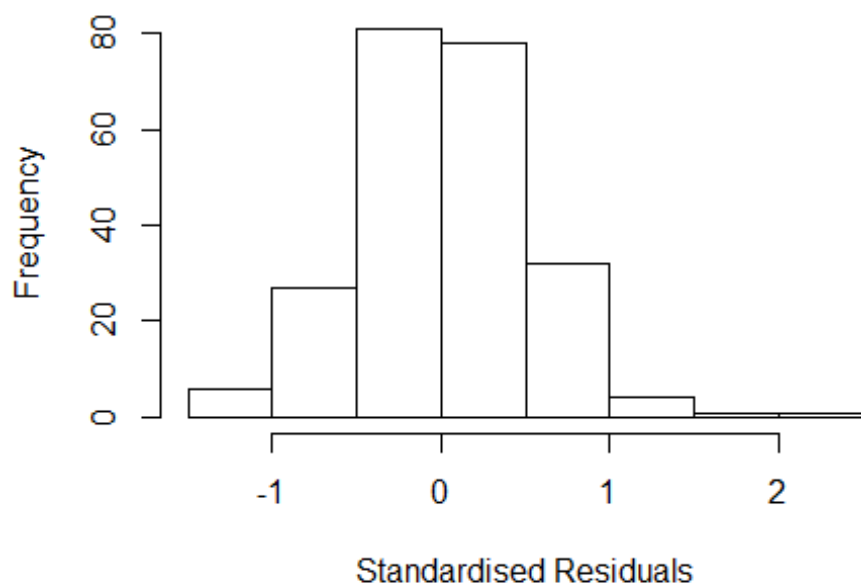


```

## Year2000          1.8138      0.0972    18.67 < 2e-16 ***
## Year2001          1.2449      0.3189     3.90 0.00013 ***
## Year2002          1.1380      0.1591     7.15 1.5e-11 ***
## Year2003          1.2242      0.1112    11.01 < 2e-16 ***
## Year2004          1.1847      0.1664     7.12 1.8e-11 ***
## Year2005          1.1220      0.1106    10.14 < 2e-16 ***
## Year2006          1.1955      0.1230     9.72 < 2e-16 ***
## Year2007          1.3308      0.1837     7.24 8.5e-12 ***
## Year2008          0.8872      0.1500     5.91 1.4e-08 ***
## Year2009          1.1055      0.1159     9.54 < 2e-16 ***
## Year2010          1.3203      0.1457     9.06 < 2e-16 ***
## Year2011          1.2920      0.1203    10.74 < 2e-16 ***
## Year2012          1.3316      0.1303    10.22 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.479
## Multiple R-squared:  0.215, Adjusted R-squared:  0.132
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 205 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0237 0.8740 0.9550 0.9050 0.9860 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.35e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##              GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 7.413e+11 1      8.610e+05
## LastAuthorFemale  3.136e+00 1      1.771e+00
## Year              2.124e+12 16      2.428e+00

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.21919 -0.31574 0.00412 0.34124 2.02891
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.83e-15 7.81e-08 0.00 1.00000
## FirstAuthorFemale1 -5.90e-02 7.46e-02 -0.79 0.42999
## LastAuthorFemale1 1.85e-01 7.61e-02 2.43 0.01583 *
## Year1997 1.12e+00 1.95e-02 57.34 < 2e-16 ***
## Year1998 1.21e+00 5.11e-01 2.37 0.01849 *
## Year1999 1.19e+00 7.46e-02 15.94 < 2e-16 ***
## Year2000 1.43e+00 1.14e-01 12.57 < 2e-16 ***
## Year2001 1.05e+00 2.79e-01 3.77 0.00021 ***
## Year2002 9.28e-01 1.47e-01 6.32 1.6e-09 ***
## Year2003 1.02e+00 1.09e-01 9.38 < 2e-16 ***
## Year2004 9.14e-01 1.48e-01 6.18 3.3e-09 ***
## Year2005 8.48e-01 1.10e-01 7.69 5.6e-13 ***
```

```

## Year2006      8.79e-01  1.11e-01  7.95  1.1e-13 ***
## Year2007      1.09e+00  1.71e-01  6.40  9.7e-10 ***
## Year2008      6.36e-01  1.48e-01  4.31  2.6e-05 ***
## Year2009      9.37e-01  1.02e-01  9.16  < 2e-16 ***
## Year2010      9.94e-01  1.20e-01  8.25  1.7e-14 ***
## Year2011      8.75e-01  9.38e-02  9.33  < 2e-16 ***
## Year2012      1.06e+00  1.19e-01  8.86  3.4e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.516
## Multiple R-squared:  0.103, Adjusted R-squared:  0.0267
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 204 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0872 0.8780 0.9540 0.9070 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.35e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 4.452e+12 1      2.110e+06
## Year      4.452e+12 16      2.485e+00

## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max

```

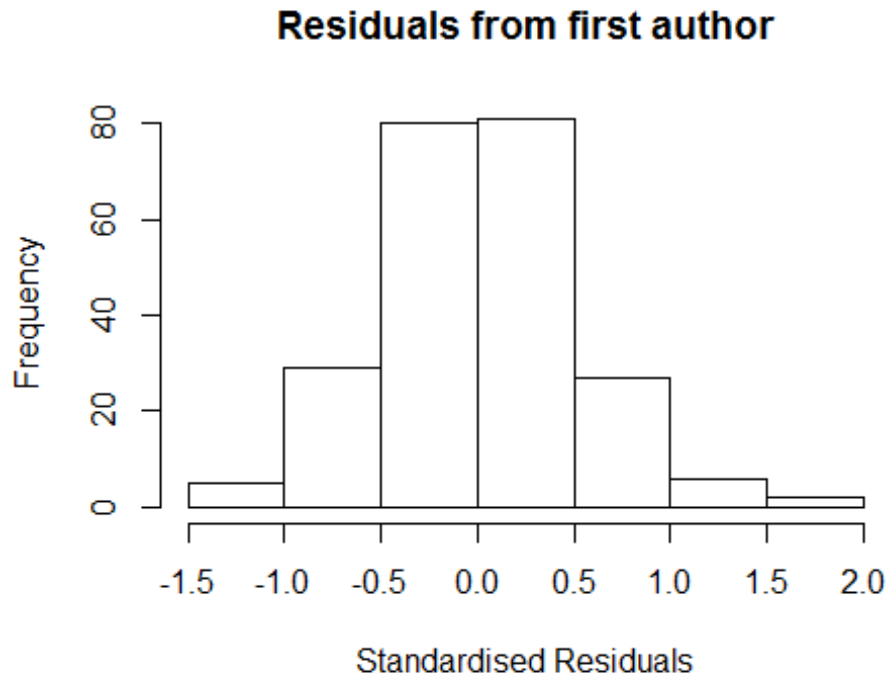
```

## -1.1682 -0.3249 0.0171 0.3247 1.9746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.41e-15  7.05e-08   0.00   1.0000
## FirstAuthorFemale1 2.08e-02  7.36e-02   0.28   0.7775
## Year1997      1.12e+00  1.95e-02  57.34 < 2e-16 ***
## Year1998      1.27e+00  4.36e-01   2.90   0.0041 **
## Year1999      1.11e+00  7.36e-02  15.09 < 2e-16 ***
## Year2000      1.45e+00  1.26e-01  11.56 < 2e-16 ***
## Year2001      1.21e+00  2.86e-01   4.23  3.5e-05 ***
## Year2002      9.38e-01  1.51e-01   6.22  2.5e-09 ***
## Year2003      1.06e+00  1.13e-01   9.37 < 2e-16 ***
## Year2004      9.02e-01  1.61e-01   5.60  6.6e-08 ***
## Year2005      9.27e-01  1.05e-01   8.81  4.5e-16 ***
## Year2006      9.32e-01  1.07e-01   8.70  9.4e-16 ***
## Year2007      1.15e+00  1.69e-01   6.80  1.1e-10 ***
## Year2008      6.89e-01  1.50e-01   4.61  7.0e-06 ***
## Year2009      9.91e-01  9.98e-02   9.93 < 2e-16 ***
## Year2010      1.09e+00  1.11e-01   9.78 < 2e-16 ***
## Year2011      9.38e-01  9.04e-02  10.38 < 2e-16 ***
## Year2012      1.12e+00  1.16e-01   9.64 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.521
## Multiple R-squared:  0.0818, Adjusted R-squared:  0.00822
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 215 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.119 0.880 0.960 0.911 0.986 0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.35e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"

```

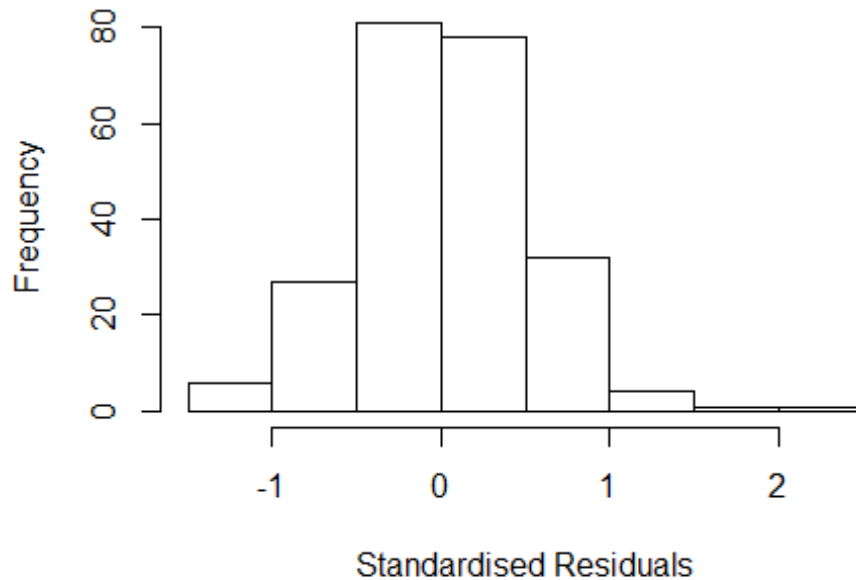
```
## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```



```
##          GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale  NaN  1          NaN
## Year              NaN 16          NaN
```

## Residuals from last author



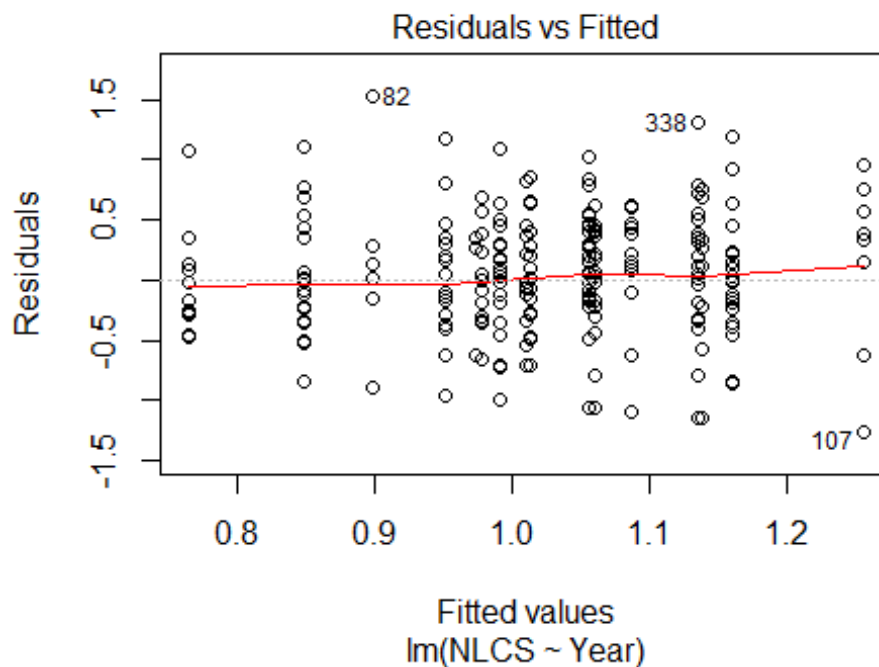
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2268 -0.3206 0.0104 0.3469 2.0538
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.99e-15 0.00e+00 Inf < 2e-16 ***
## LastAuthorFemale1 1.59e-01 7.37e-02 2.15 0.03257 *
## Year1997 1.12e+00 1.95e-02 57.34 < 2e-16 ***
## Year1998 1.20e+00 5.32e-01 2.25 0.02563 *
## Year1999 1.13e+00 0.00e+00 Inf < 2e-16 ***
## Year2000 1.42e+00 1.06e-01 13.37 < 2e-16 ***
## Year2001 1.06e+00 2.89e-01 3.65 0.00033 ***
## Year2002 8.99e-01 1.42e-01 6.35 1.3e-09 ***
## Year2003 1.00e+00 1.03e-01 9.70 < 2e-16 ***
## Year2004 8.94e-01 1.48e-01 6.04 6.9e-09 ***
## Year2005 8.34e-01 1.05e-01 7.91 1.4e-13 ***
## Year2006 8.55e-01 1.02e-01 8.37 8.0e-15 ***
```

```

## Year2007          1.07e+00  1.60e-01  6.69  1.9e-10 ***
## Year2008          6.20e-01  1.46e-01  4.25  3.3e-05 ***
## Year2009          9.21e-01  1.01e-01  9.16  < 2e-16 ***
## Year2010          9.80e-01  1.22e-01  8.02  6.9e-14 ***
## Year2011          8.60e-01  9.31e-02  9.24  < 2e-16 ***
## Year2012          1.05e+00  1.20e-01  8.71  8.9e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.101, Adjusted R-squared:  0.0288
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 26 weights are ~= 1. The remaining 204 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0825 0.8740 0.9560 0.9070 0.9840 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.35e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 230"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3607"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   12   21    8   16   18   19   22   17   20   21   21   19   26   25   31
## 2011 2012
##   39   27
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    9    8    3   11    7    9   14   15   13   16   17   12   21   20   22
## 2011 2012

```

```
## 32 21
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 9 7 2 10 7 9 14 14 13 13 15 11 20 20 19
## 2011 2012
## 31 20
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```

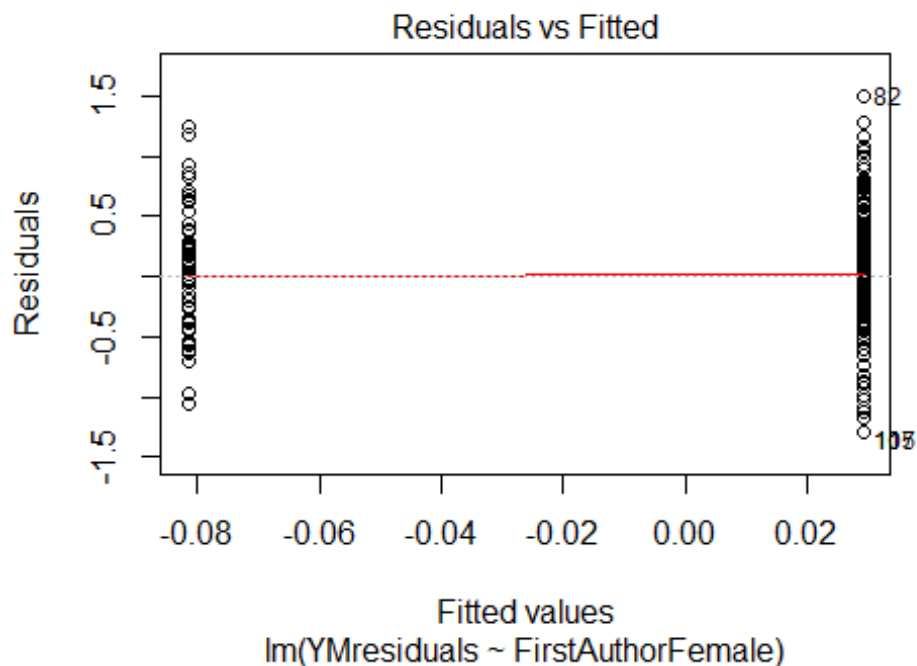


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.0069, df = 1, p-value = 0.9
## [1] "Female first author team size 2018 geometric mean: 3.55655882007785"
## [1] "Male first author team size 2018 geometric mean: 1.8612097182042"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.93242414866094"
## [1] "Male last author team size 2018 geometric mean: 1.81406355387873"

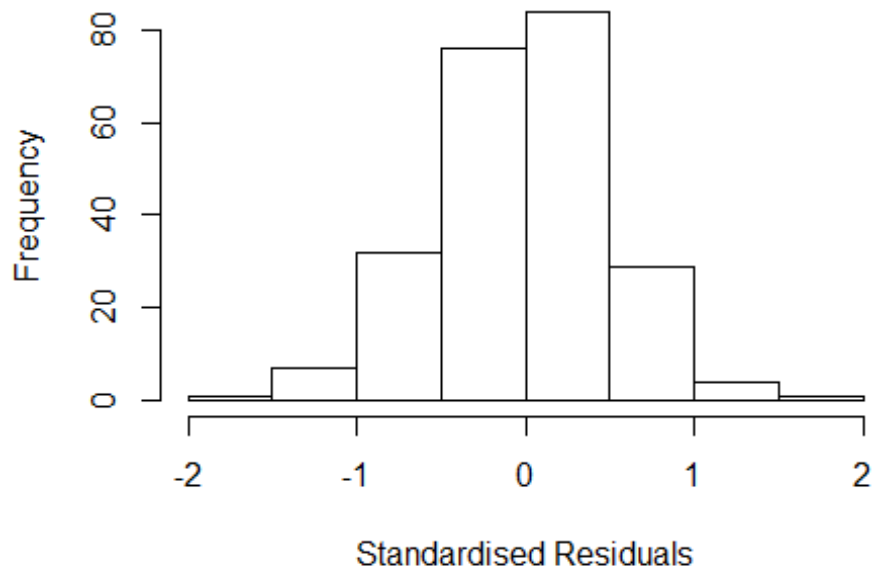
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 38, p-value = 0.01
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
## FirstAuthorFemale	1.880	1	1.371
## LastAuthorFemale	1.497	1	1.224
## UniqueAuthors	4.537	4	1.208
## Year	7.731	16	1.066

## Residuals from first and last author and team size



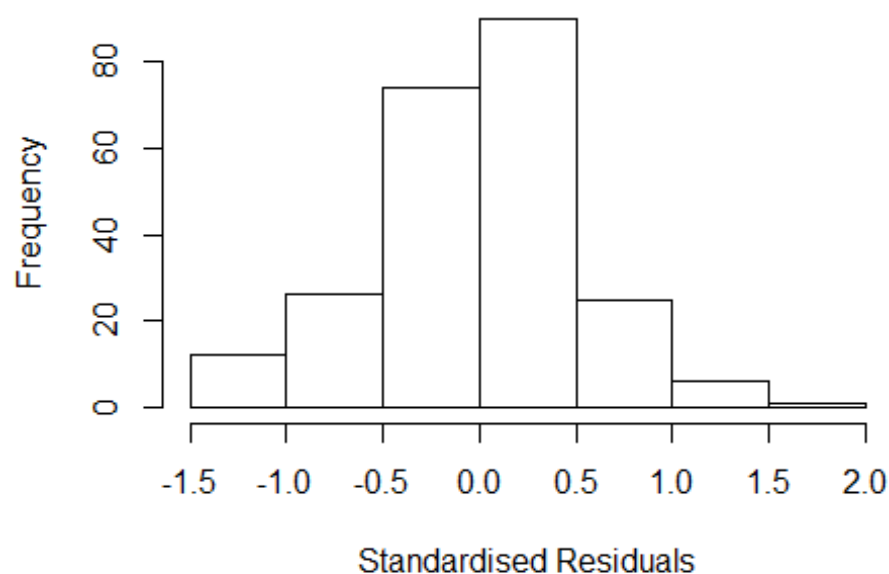
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.6162 -0.3196 0.0234 0.3480 1.5347
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.90755 0.28749 3.16 0.00183 **
## FirstAuthorFemale1 -0.14508 0.09647 -1.50 0.13412
## LastAuthorFemale1 -0.06032 0.09550 -0.63 0.52835
## UniqueAuthors2 0.30453 0.12273 2.48 0.01387 *
## UniqueAuthors3 0.29258 0.12996 2.25 0.02539 *
## UniqueAuthors4 0.41146 0.10732 3.83 0.00017 ***
## UniqueAuthors5 0.26882 0.10233 2.63 0.00925 **
## Year1997 0.10400 0.33270 0.31 0.75490
## Year1998 -0.02872 0.38616 -0.07 0.94077
## Year1999 -0.16986 0.28926 -0.59 0.55768
```

```

## Year2000      -0.32277      0.35296      -0.91      0.36151
## Year2001      0.41607      0.46706      0.89      0.37404
## Year2002     -0.00451      0.28781     -0.02      0.98751
## Year2003      0.03313      0.30888      0.11      0.91469
## Year2004     -0.32646      0.28834     -1.13      0.25883
## Year2005      0.04195      0.32883      0.13      0.89861
## Year2006      0.11481      0.30122      0.38      0.70348
## Year2007     -0.06449      0.29966     -0.22      0.82982
## Year2008     -0.09352      0.28244     -0.33      0.74089
## Year2009     -0.14912      0.29357     -0.51      0.61203
## Year2010     -0.00748      0.30093     -0.02      0.98019
## Year2011     -0.04163      0.28438     -0.15      0.88377
## Year2012     -0.17160      0.29363     -0.58      0.55957
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.514
## Multiple R-squared:  0.146, Adjusted R-squared:  0.0574
## Convergence in 19 IRWLS iterations
##
## Robustness weights:
## 11 weights are ~= 1. The remaining 223 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.302  0.892  0.955   0.911   0.988   0.999
## Algorithmic parameters:
##           tuning.chi           bb           tuning.psi           refine.tol
##           1.55e+00           5.00e-01           4.69e+00           1.00e-07
##           rel.tol           solve.tol           eps.outlier           eps.x
##           1.00e-07           1.00e-07           4.27e-04           1.82e-12
## warn.limit.reject warn.limit.meanrw
##           5.00e-01           5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##           500           50           2           1           1000           200
## trace.lev      mts      compute.rd
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.798 1           1.341
## LastAuthorFemale  1.318 1           1.148
## Year              1.911 16           1.020

```

## Residuals from first and last author



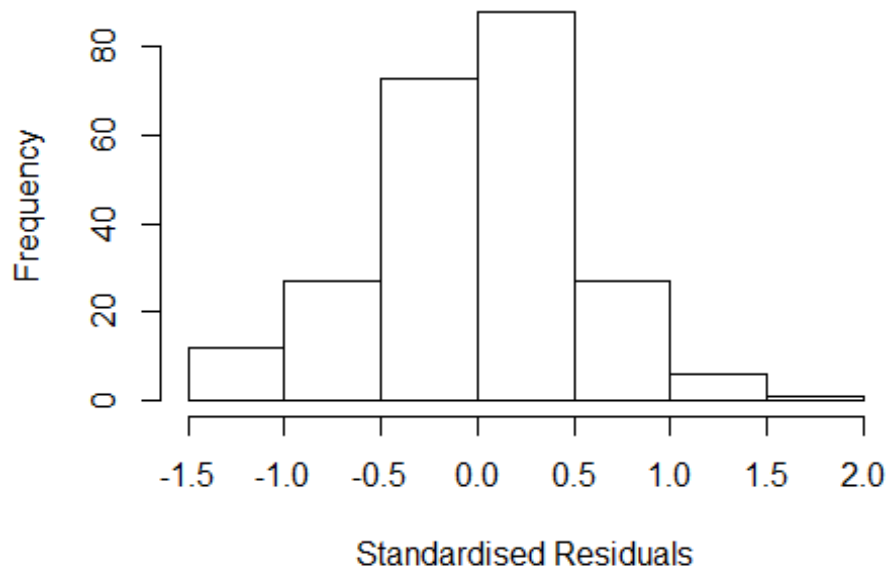
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4850 -0.3257 0.0248 0.3383 1.6653
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.2095 0.2124 5.69 4e-08 ***
## FirstAuthorFemale1 -0.1220 0.0987 -1.24 0.22
## LastAuthorFemale1 -0.0741 0.0983 -0.75 0.45
## Year1997 0.0165 0.2934 0.06 0.96
## Year1998 -0.2125 0.4478 -0.47 0.64
## Year1999 -0.2168 0.2495 -0.87 0.39
## Year2000 -0.4508 0.3517 -1.28 0.20
## Year2001 0.2755 0.3998 0.69 0.49
## Year2002 -0.0631 0.2477 -0.25 0.80
## Year2003 -0.0264 0.2684 -0.10 0.92
## Year2004 -0.4362 0.2303 -1.89 0.06 .
## Year2005 -0.1868 0.2741 -0.68 0.50
```

```

## Year2006          -0.0800      0.2371   -0.34      0.74
## Year2007          -0.2022      0.2465   -0.82      0.41
## Year2008          -0.1488      0.2418   -0.62      0.54
## Year2009          -0.2077      0.2490   -0.83      0.41
## Year2010          -0.0751      0.2668   -0.28      0.78
## Year2011          -0.1036      0.2360   -0.44      0.66
## Year2012          -0.3236      0.2461   -1.31      0.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.504
## Multiple R-squared:  0.0862, Adjusted R-squared:  0.00971
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 18 weights are ~= 1. The remaining 216 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.253  0.860  0.954  0.898  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      4.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.597 1      1.264
## Year              1.597 16      1.015

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4872 -0.3217 0.0271 0.3441 1.6673
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18094 0.19995 5.91 1.3e-08 ***
## FirstAuthorFemale1 -0.14490 0.09368 -1.55 0.12
## Year1997 0.05376 0.27939 0.19 0.85
## Year1998 -0.19804 0.48383 -0.41 0.68
## Year1999 -0.18611 0.23519 -0.79 0.43
## Year2000 -0.42426 0.34448 -1.23 0.22
## Year2001 0.30626 0.38182 0.80 0.42
## Year2002 -0.04466 0.24020 -0.19 0.85
## Year2003 -0.00968 0.26198 -0.04 0.97
## Year2004 -0.42312 0.22365 -1.89 0.06 .
## Year2005 -0.16117 0.26587 -0.61 0.55
## Year2006 -0.05596 0.22864 -0.24 0.81
```

```

## Year2007          -0.18302    0.23927   -0.76    0.45
## Year2008          -0.13240    0.23471   -0.56    0.57
## Year2009          -0.18073    0.23900   -0.76    0.45
## Year2010          -0.06381    0.26123   -0.24    0.81
## Year2011          -0.08123    0.22753   -0.36    0.72
## Year2012          -0.30797    0.24039   -1.28    0.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.085, Adjusted R-squared:  0.013
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 19 weights are ~= 1. The remaining 215 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.245  0.856  0.952   0.896  0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.181 1          1.087
## Year            1.181 16          1.005
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.4615 -0.3377  0.0204  0.3483  1.6674

```

```

##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.2303    0.2165    5.68 4.3e-08 ***
## LastAuthorFemale1 -0.1171    0.0953   -1.23  0.220
## Year1997          -0.0482    0.2864   -0.17  0.866
## Year1998          -0.3338    0.4141   -0.81  0.421
## Year1999          -0.2855    0.2461   -1.16  0.247
## Year2000          -0.4737    0.3538   -1.34  0.182
## Year2001           0.2311    0.4135    0.56  0.577
## Year2002          -0.0996    0.2477   -0.40  0.688
## Year2003          -0.0800    0.2642   -0.30  0.762
## Year2004          -0.4666    0.2346   -1.99  0.048 *
## Year2005          -0.2291    0.2736   -0.84  0.403
## Year2006          -0.1077    0.2424   -0.44  0.657
## Year2007          -0.2374    0.2487   -0.95  0.341
## Year2008          -0.2014    0.2402   -0.84  0.403
## Year2009          -0.2781    0.2365   -1.18  0.241
## Year2010          -0.1139    0.2684   -0.42  0.672
## Year2011          -0.1341    0.2391   -0.56  0.576
## Year2012          -0.3710    0.2419   -1.53  0.127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.501
## Multiple R-squared:  0.0799, Adjusted R-squared:  0.00752
## Convergence in 17 IRWLS iterations
##
## Robustness weights:
## 20 weights are ~= 1. The remaining 214 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.245  0.864  0.952  0.895  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      4.27e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 234"
## [1] ""

```



```

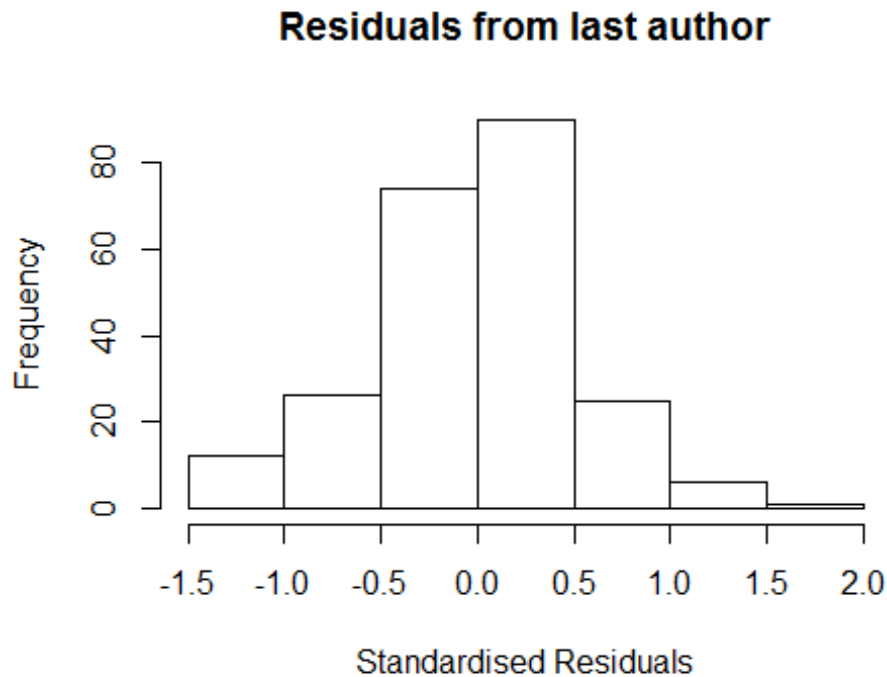
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3608"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
## < table of extent 0 >
## < table of extent 0 >
## < table of extent 0 >
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 1.44224957030741"
## [1] "Male first author team size 2018 geometric mean: 1"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 1.44224957030741"
## [1] "Male last author team size 2018 geometric mean: 1"

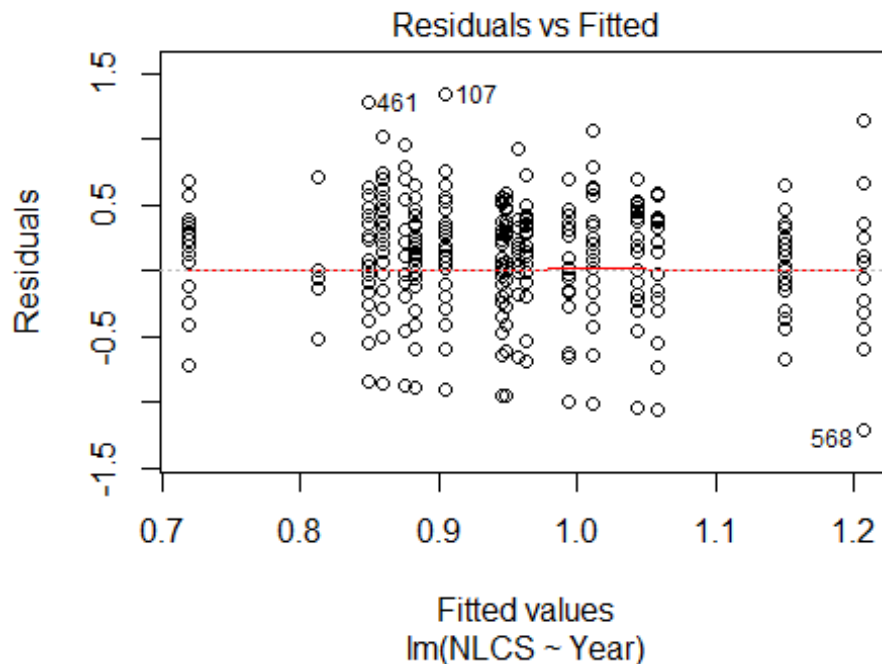
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 2, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
## [1] "Regression 3: First author gender, Year as factors"
## [1] "Regression 4: Last author gender, Year as factors"
## [1] "Sample size for the above analysis: 0"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3609"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 41 44 54 22 43 42 31 37 31 39 37 48 41 19 39
## 2011 2012
## 35 31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 34 36 49 20 16 5 24 31 25 36 30 39 38 13 29
```

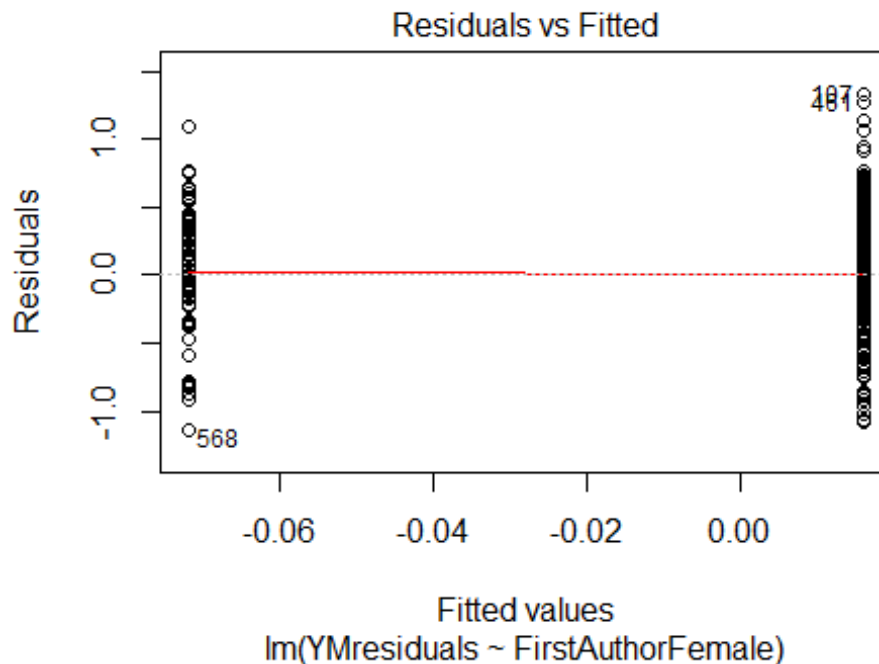
```
## 2011 2012
## 33 22
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 33 36 49 18 15 5 23 27 22 32 29 38 36 11 25
## 2011 2012
## 30 20
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 15, df = 16, p-value = 0.5
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.11, df = 1, p-value = 0.7
## [1] "Female first author team size 2018 geometric mean: 3.50213833521581"
## [1] "Male first author team size 2018 geometric mean: 1"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 22, p-value = 0.03
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.35884046647934"
## [1] "Male last author team size 2018 geometric mean: 2.26793315526605"

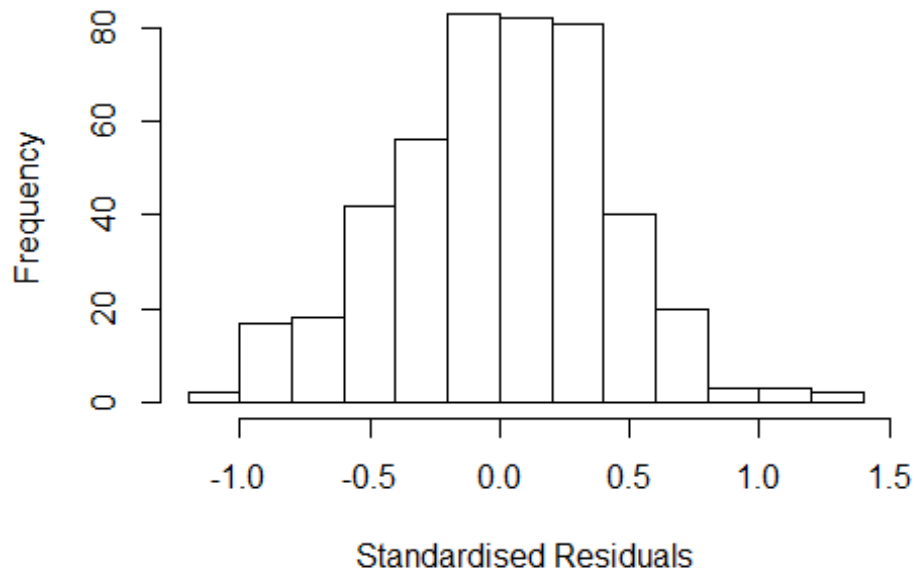
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 26, p-value = 0.5
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.384	1	1.176
LastAuthorFemale	1.331	1	1.154
UniqueAuthors	2.335	4	1.112
Year	2.746	16	1.032

## Residuals from first and last author and team size



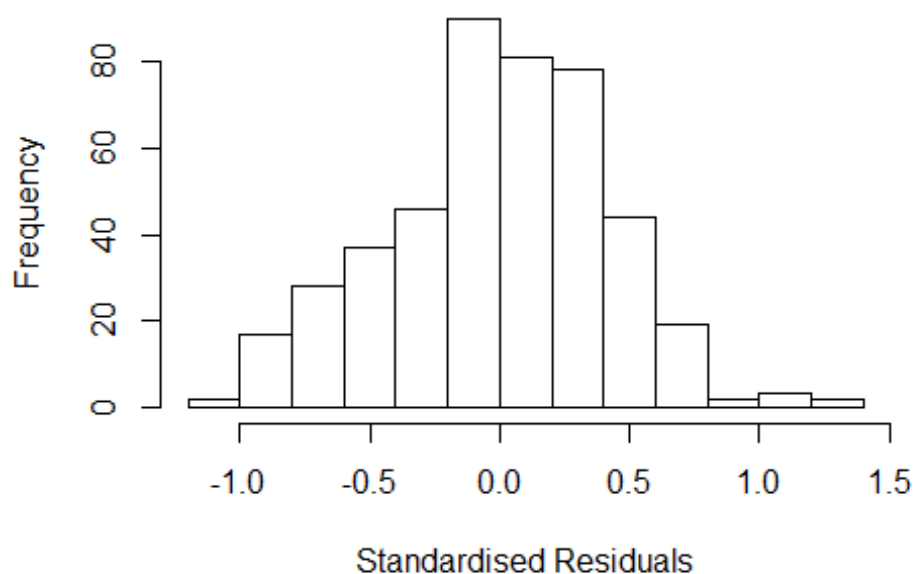
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
##       UniqueAuthors +
##       Year, data = AllScopusDataOlderFirstLastGendered, control =
##       lmrob.control(fast.s.large.n = Inf,
##       k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.09485 -0.27226  0.00791  0.28931  1.35591
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94099    0.09223   10.20  <2e-16 ***
## FirstAuthorFemale1  0.05541    0.06006    0.92   0.357
## LastAuthorFemale1 -0.00456    0.05181   -0.09   0.930
## UniqueAuthors2     0.00238    0.04837    0.05   0.961
## UniqueAuthors3     0.03053    0.07233    0.42   0.673
## UniqueAuthors4     0.10820    0.10601    1.02   0.308
## UniqueAuthors5     0.17810    0.09698    1.84   0.067 .
## Year1997         -0.08377    0.13461   -0.62   0.534
## Year1998         -0.09974    0.10006   -1.00   0.319
## Year1999          0.16340    0.11431    1.43   0.154
```

```

## Year2000      -0.01805    0.13440   -0.13    0.893
## Year2001      -0.19351    0.20056   -0.96    0.335
## Year2002      -0.04492    0.12222   -0.37    0.713
## Year2003       0.02135    0.10987    0.19    0.846
## Year2004       0.10064    0.12920    0.78    0.436
## Year2005      -0.11677    0.10510   -1.11    0.267
## Year2006      -0.24743    0.11229   -2.20    0.028 *
## Year2007      -0.15236    0.09642   -1.58    0.115
## Year2008      -0.06951    0.10721   -0.65    0.517
## Year2009       0.17617    0.16758    1.05    0.294
## Year2010      -0.05168    0.13201   -0.39    0.696
## Year2011       0.04388    0.10908    0.40    0.688
## Year2012      -0.20166    0.15469   -1.30    0.193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.413
## Multiple R-squared:  0.0684, Adjusted R-squared:  0.0203
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 41 weights are ~= 1. The remaining 408 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.259  0.858  0.950  0.905  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.348 1      1.161
## LastAuthorFemale  1.264 1      1.124
## Year              1.266 16      1.007

```

## Residuals from first and last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0970 -0.2705 0.0119 0.2999 1.3545
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.94997 0.08864 10.72 <2e-16 ***
## FirstAuthorFemale1 0.06207 0.05971 1.04 0.299
## LastAuthorFemale1 -0.01504 0.05140 -0.29 0.770
## Year1997 -0.08707 0.13437 -0.65 0.517
## Year1998 -0.10345 0.09904 -1.04 0.297
## Year1999 0.15919 0.11347 1.40 0.161
## Year2000 -0.01948 0.13428 -0.15 0.885
## Year2001 -0.20013 0.19902 -1.01 0.315
## Year2002 -0.02881 0.12235 -0.24 0.814
## Year2003 0.01958 0.11021 0.18 0.859
## Year2004 0.10004 0.13004 0.77 0.442
## Year2005 -0.09767 0.10541 -0.93 0.355
```

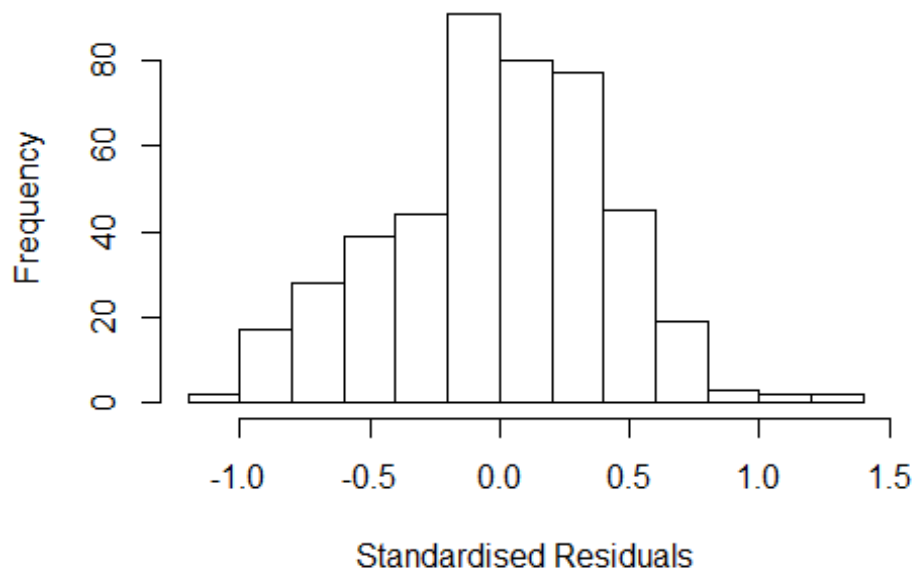
```

## Year2006      -0.23788    0.11252   -2.11    0.035 *
## Year2007      -0.13783    0.09805   -1.41    0.161
## Year2008      -0.05790    0.11110   -0.52    0.603
## Year2009       0.21195    0.15279    1.39    0.166
## Year2010       0.00148    0.12795    0.01    0.991
## Year2011       0.06942    0.10727    0.65    0.518
## Year2012      -0.16468    0.14652   -1.12    0.262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.41
## Multiple R-squared:  0.062, Adjusted R-squared:  0.0228
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 411 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.252  0.863   0.948   0.903   0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50           2           1           1000      200
##      trace.lev      mts      compute.rd
##      0             1000      0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.154 1      1.074
## Year              1.154 16      1.004

```



## Residuals from first author



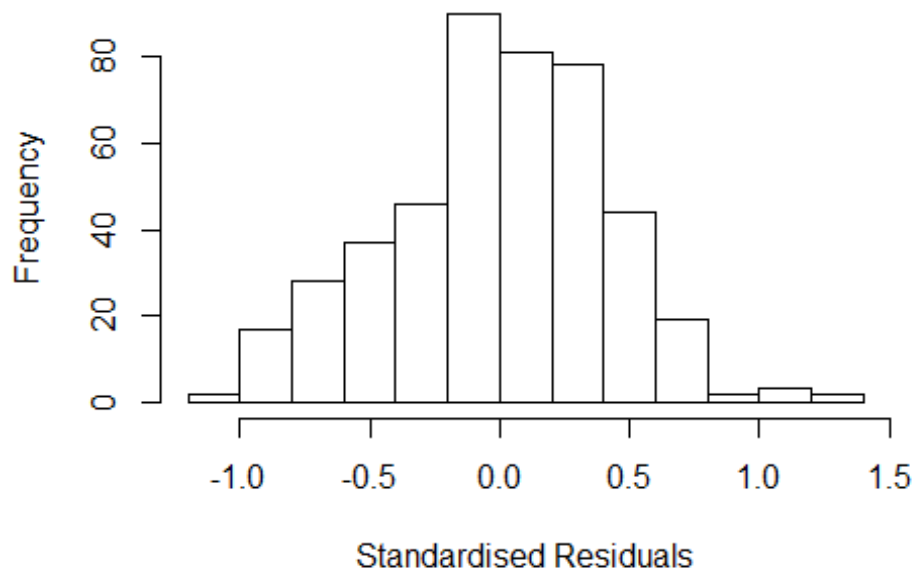
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0984 -0.2704  0.0139  0.3020  1.3515
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.94477    0.08712   10.84  <2e-16 ***
## FirstAuthorFemale1 0.05476    0.05534    0.99   0.323
## Year1997       -0.08961    0.13392   -0.67   0.504
## Year1998       -0.10303    0.09899   -1.04   0.299
## Year1999        0.15921    0.11384    1.40   0.163
## Year2000       -0.02141    0.13417   -0.16   0.873
## Year2001       -0.20050    0.19897   -1.01   0.314
## Year2002       -0.02808    0.12224   -0.23   0.818
## Year2003        0.02052    0.11059    0.19   0.853
## Year2004        0.09889    0.12950    0.76   0.446
## Year2005       -0.09901    0.10520   -0.94   0.347
## Year2006       -0.23897    0.11209   -2.13   0.034 *
```

```

## Year2007          -0.13871    0.09779   -1.42    0.157
## Year2008          -0.05778    0.11141   -0.52    0.604
## Year2009           0.21264    0.15204    1.40    0.163
## Year2010           0.00223    0.12818    0.02    0.986
## Year2011           0.07067    0.10740    0.66    0.511
## Year2012          -0.16344    0.14638   -1.12    0.265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.411
## Multiple R-squared:  0.0618, Adjusted R-squared:  0.0248
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 40 weights are ~= 1. The remaining 409 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.257  0.864  0.947  0.904  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.09 1          1.044
## Year              1.09 16          1.003

```

## Residuals from last author



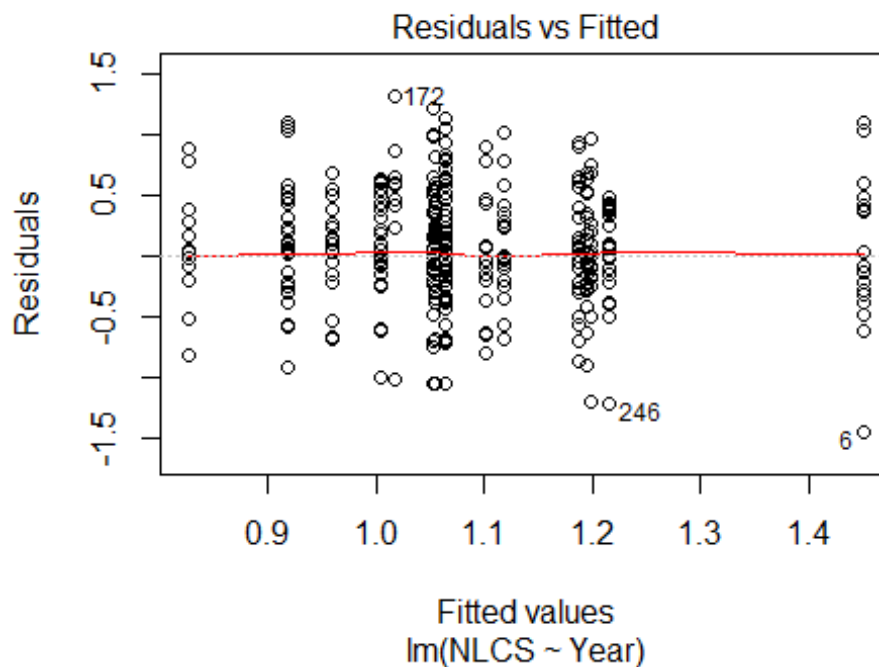
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.0885 -0.2615 0.0148 0.2945 1.3569
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.98075 0.08533 11.49 <2e-16 ***
## LastAuthorFemale1 0.00876 0.04815 0.18 0.856
## Year1997 -0.09237 0.13544 -0.68 0.496
## Year1998 -0.09844 0.09863 -1.00 0.319
## Year1999 0.16489 0.11316 1.46 0.146
## Year2000 -0.01540 0.13432 -0.11 0.909
## Year2001 -0.20271 0.19984 -1.01 0.311
## Year2002 -0.02763 0.12125 -0.23 0.820
## Year2003 0.02897 0.10928 0.27 0.791
## Year2004 0.09895 0.13099 0.76 0.450
## Year2005 -0.09316 0.10473 -0.89 0.374
## Year2006 -0.24169 0.11358 -2.13 0.034 *
```

```

## Year2007          -0.14000      0.09784    -1.43      0.153
## Year2008          -0.05329      0.11088    -0.48      0.631
## Year2009           0.21890      0.14859     1.47      0.141
## Year2010           0.01201      0.12698     0.09      0.925
## Year2011           0.07027      0.10900     0.64      0.520
## Year2012          -0.16751      0.14519    -1.15      0.249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.409
## Multiple R-squared:  0.0596, Adjusted R-squared:  0.0225
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 38 weights are ~= 1. The remaining 411 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.247  0.864   0.947   0.903   0.986   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.23e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 449"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3610"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   34   34   23   29   27   31   33   23   34   32   23   53   32   58   43
## 2011 2012
##   55   36
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   21   17   14   19   13   19   26   16   24   27   14   35   28   35   29
## 2011 2012

```

```
## 43 29
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 12 14 17 12 18 22 15 22 25 12 32 24 28 24
## 2011 2012
## 32 25
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 32, df = 16, p-value = 0.01
```



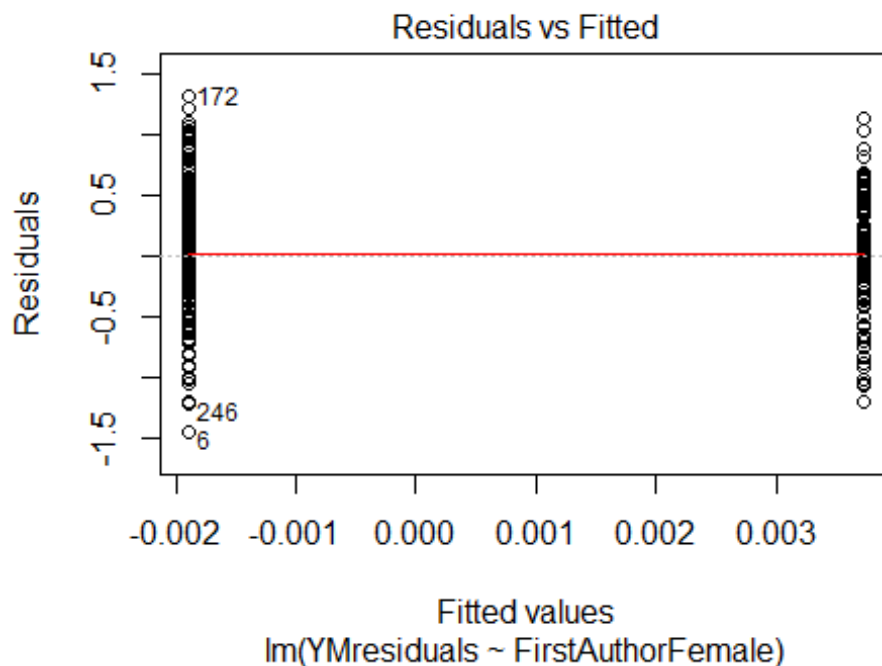
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 1.5, df = 1, p-value = 0.2

## [1] "Female first author team size 2018 geometric mean: 3.39852287976549"
## [1] "Male first author team size 2018 geometric mean: 2.39860786693316"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

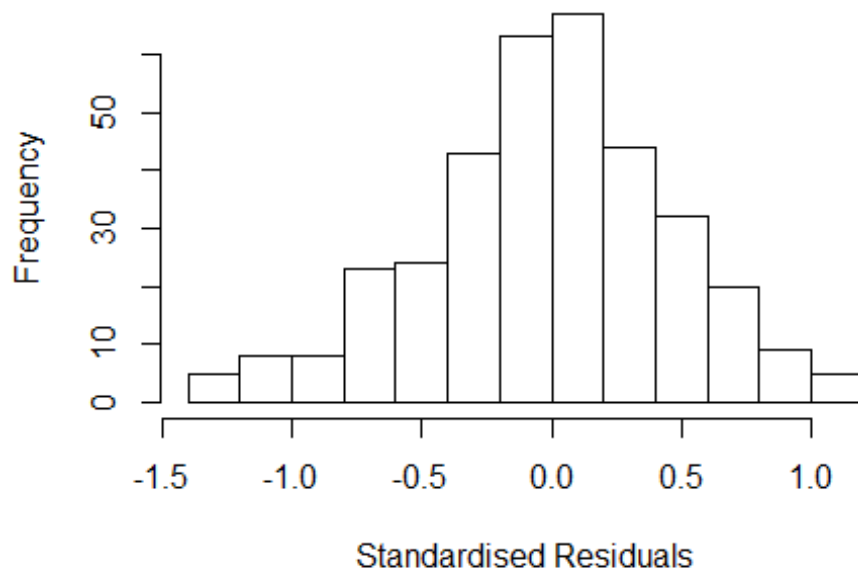
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 64, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 4.29995060728035"
## [1] "Male last author team size 2018 geometric mean: 2.43364045373981"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 58, p-value = 0.08
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.576 1      1.255
## LastAuthorFemale  1.508 1      1.228
## UniqueAuthors    3.846 4      1.183
## Year              7.216 16     1.064
```

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.26995 -0.30999 0.00864 0.28785 1.17879
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.26995 0.13150 9.66 < 2e-16 ***
## FirstAuthorFemale1 -0.04272 0.06231 -0.69 0.49345
## LastAuthorFemale1 -0.08279 0.06661 -1.24 0.21484
## UniqueAuthors2 0.25908 0.09227 2.81 0.00529 **
## UniqueAuthors3 0.34255 0.09187 3.73 0.00023 ***
## UniqueAuthors4 0.30723 0.10784 2.85 0.00466 **
## UniqueAuthors5 0.34243 0.13064 2.62 0.00917 **
## Year1997 -0.00209 0.28686 -0.01 0.99418
## Year1998 -0.60676 0.17542 -3.46 0.00061 ***
## Year1999 -0.37789 0.14996 -2.52 0.01221 *
```

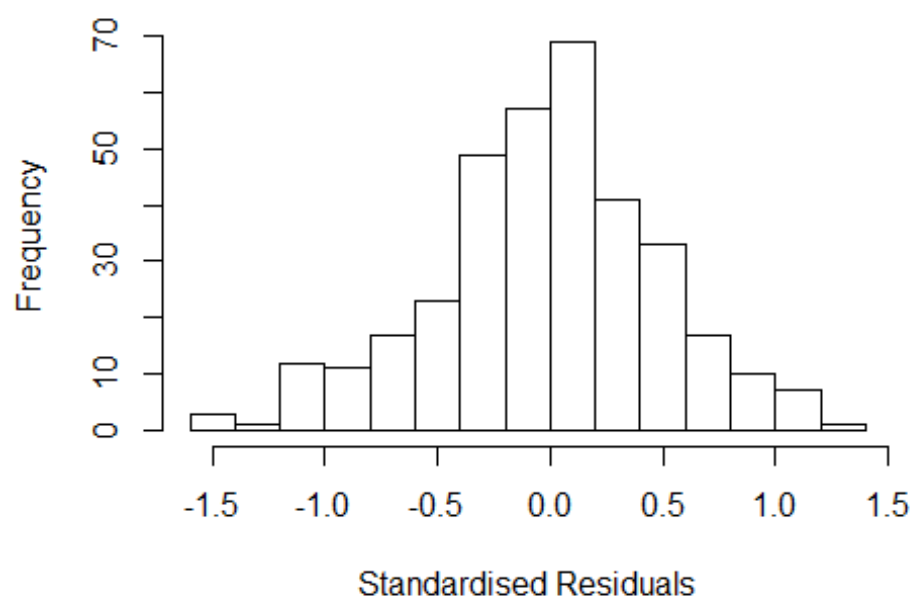
```

## Year2000      -0.38252    0.72486   -0.53   0.59806
## Year2001      -0.29775    0.17919   -1.66   0.09755 .
## Year2002      -0.27440    0.15845   -1.73   0.08425 .
## Year2003      -0.36973    0.18726   -1.97   0.04918 *
## Year2004      -0.41034    0.15147   -2.71   0.00710 **
## Year2005      -0.31385    0.14686   -2.14   0.03333 *
## Year2006      -0.40999    0.19644   -2.09   0.03765 *
## Year2007      -0.54101    0.14976   -3.61   0.00035 ***
## Year2008      -0.28681    0.16113   -1.78   0.07601 .
## Year2009      -0.40434    0.15821   -2.56   0.01105 *
## Year2010      -0.51885    0.17167   -3.02   0.00271 **
## Year2011      -0.53167    0.14665   -3.63   0.00033 ***
## Year2012      -0.41227    0.16063   -2.57   0.01071 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.446
## Multiple R-squared:  0.131, Adjusted R-squared:  0.0728
## Convergence in 39 IRWLS iterations
##
## Robustness weights:
## 34 weights are ~= 1. The remaining 317 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.398  0.852  0.948   0.893  0.987   0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.85e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.748 1 1.322
## LastAuthorFemale 1.553 1 1.246
## Year 2.492 16 1.029

```



## Residuals from first and last author



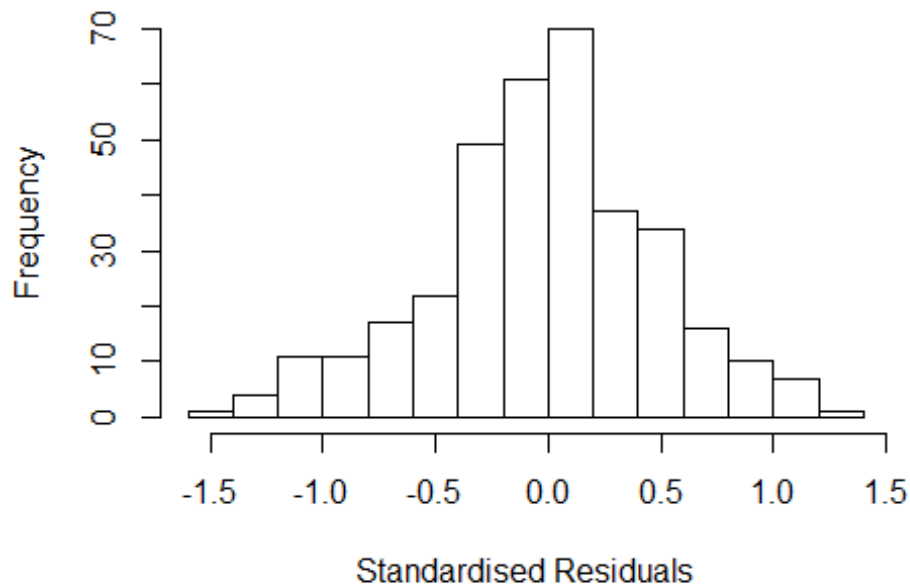
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47288 -0.27526  0.00132  0.29614  1.23746
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.472880   0.130908   11.25 < 2e-16 ***
## FirstAuthorFemale1  0.000342   0.065726    0.01  0.99586
## LastAuthorFemale1 -0.075271   0.069443   -1.08  0.27918
## Year1997         -0.051433   0.302485   -0.17  0.86509
## Year1998         -0.643709   0.189230   -3.40  0.00075 ***
## Year1999         -0.430768   0.164088   -2.63  0.00906 **
## Year2000         -0.385340   0.760150   -0.51  0.61254
## Year2001         -0.236672   0.175963   -1.35  0.17954
## Year2002         -0.201251   0.160332   -1.26  0.21028
## Year2003         -0.357647   0.197145   -1.81  0.07056 .
## Year2004         -0.375544   0.161935   -2.32  0.02100 *
## Year2005         -0.238617   0.150736   -1.58  0.11437
```

```

## Year2006          -0.356771    0.207015    -1.72    0.08575  .
## Year2007          -0.539068    0.161003    -3.35    0.00091 ***
## Year2008          -0.262610    0.159667    -1.64    0.10097
## Year2009          -0.331164    0.158709    -2.09    0.03769 *
## Year2010          -0.472601    0.175732    -2.69    0.00752 **
## Year2011          -0.435323    0.146825    -2.96    0.00325 **
## Year2012          -0.381598    0.164937    -2.31    0.02130 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.438
## Multiple R-squared:  0.0872, Adjusted R-squared:  0.0377
## Convergence in 44 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 314 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.    Max.
##  0.234  0.836  0.949  0.882  0.984  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##   trace.lev      mts      compute.rd
##      0          1000          0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.808 1          1.345
## Year              1.808 16          1.019

```

## Residuals from first author



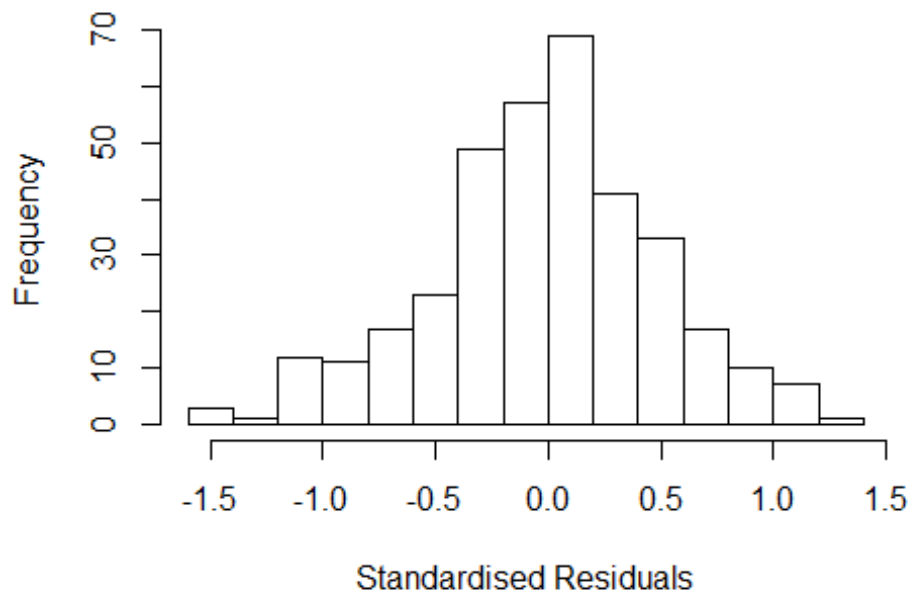
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4675 -0.2745 -0.0107 0.3107 1.2283
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.46753 0.13095 11.21 < 2e-16 ***
## FirstAuthorFemale1 -0.00999 0.06681 -0.15 0.88127
## Year1997 -0.08554 0.29389 -0.29 0.77118
## Year1998 -0.63611 0.18899 -3.37 0.00085 ***
## Year1999 -0.43839 0.16467 -2.66 0.00814 **
## Year2000 -0.37085 0.79737 -0.47 0.64217
## Year2001 -0.26169 0.17682 -1.48 0.13982
## Year2002 -0.20990 0.16236 -1.29 0.19697
## Year2003 -0.37005 0.19691 -1.88 0.06109 .
## Year2004 -0.37786 0.16338 -2.31 0.02134 *
## Year2005 -0.25181 0.15022 -1.68 0.09463 .
## Year2006 -0.38132 0.20600 -1.85 0.06504 .
```

```

## Year2007          -0.54761    0.16271   -3.37  0.00085 ***
## Year2008          -0.27160    0.15986   -1.70  0.09025 .
## Year2009          -0.34386    0.15798   -2.18  0.03021 *
## Year2010          -0.47659    0.17794   -2.68  0.00777 **
## Year2011          -0.43048    0.14664   -2.94  0.00356 **
## Year2012          -0.40515    0.16158   -2.51  0.01264 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.436
## Multiple R-squared:  0.0816, Adjusted R-squared:  0.0347
## Convergence in 47 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.234  0.836  0.952  0.881  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.565 1          1.251
## Year            1.565 16          1.014

```

## Residuals from last author



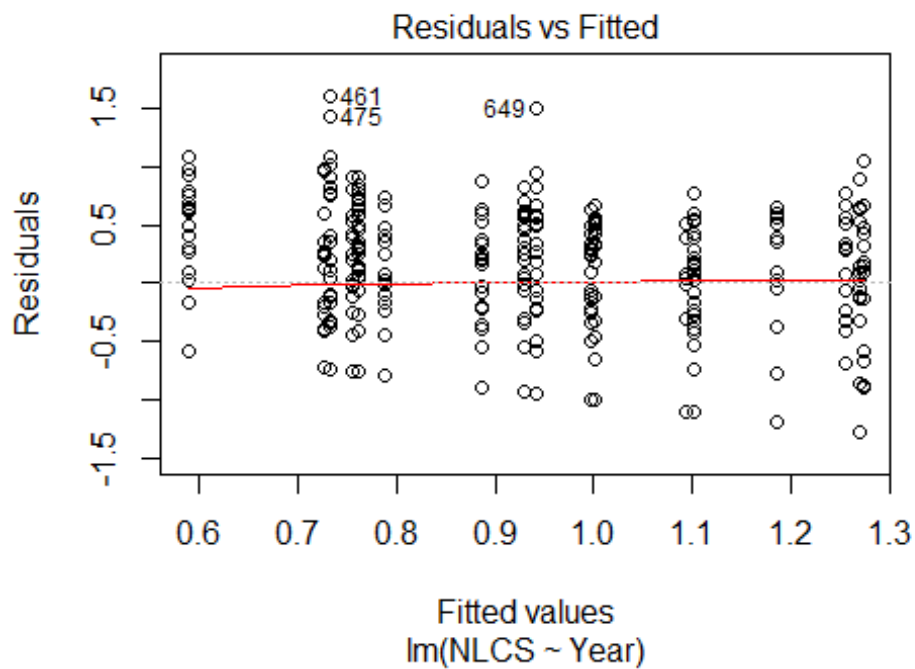
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.47292 -0.27528 0.00165 0.29617 1.23549
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.4729 0.1305 11.29 < 2e-16 ***
## LastAuthorFemale1 -0.0752 0.0698 -1.08 0.28192
## Year1997 -0.0502 0.2837 -0.18 0.85975
## Year1998 -0.6437 0.1892 -3.40 0.00075 ***
## Year1999 -0.4306 0.1640 -2.63 0.00905 **
## Year2000 -0.3834 0.7902 -0.49 0.62786
## Year2001 -0.2365 0.1759 -1.34 0.17975
## Year2002 -0.2011 0.1603 -1.25 0.21046
## Year2003 -0.3576 0.1966 -1.82 0.06977 .
## Year2004 -0.3756 0.1605 -2.34 0.01987 *
## Year2005 -0.2385 0.1505 -1.58 0.11405
## Year2006 -0.3566 0.2059 -1.73 0.08429 .
```

```

## Year2007          -0.5391      0.1613    -3.34  0.00093 ***
## Year2008          -0.2626      0.1596    -1.65  0.10083
## Year2009          -0.3311      0.1559    -2.12  0.03442 *
## Year2010          -0.4724      0.1739    -2.72  0.00693 **
## Year2011          -0.4352      0.1456    -2.99  0.00300 **
## Year2012          -0.3813      0.1630    -2.34  0.01987 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.437
## Multiple R-squared:  0.0874, Adjusted R-squared:  0.0408
## Convergence in 41 IRWLS iterations
##
## Robustness weights:
## 37 weights are ~= 1. The remaining 314 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  0.835   0.949   0.881   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.85e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 351"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3611"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   33   28   26   26   40   32   28   30   22   22   20   40   74   48   41
## 2011 2012
##   55   34
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   18   15   16   18   18    8   21   25   20   16   18   31   45   42   32
## 2011 2012

```

```
## 40 31
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 17 14 16 17 17 8 18 23 20 12 16 30 42 42 27
## 2011 2012
## 35 31
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 14, df = 16, p-value = 0.6
```



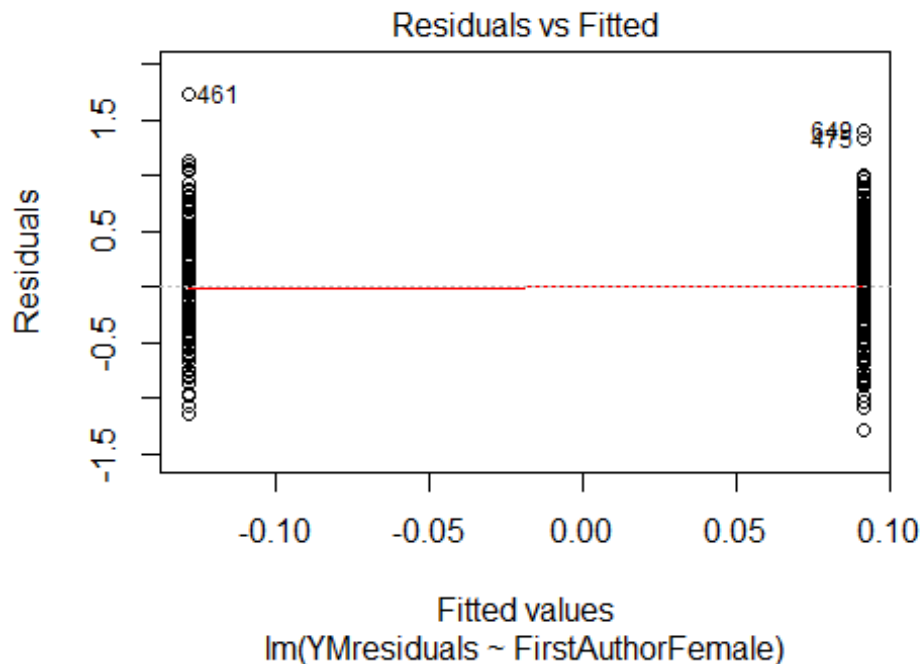
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 0.016, df = 1, p-value = 0.9

## [1] "Female first author team size 2018 geometric mean: 2.81342699994754"
## [1] "Male first author team size 2018 geometric mean: 3.65310315919376"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 180, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.00178234218487"
## [1] "Male last author team size 2018 geometric mean: 3.43697676676045"

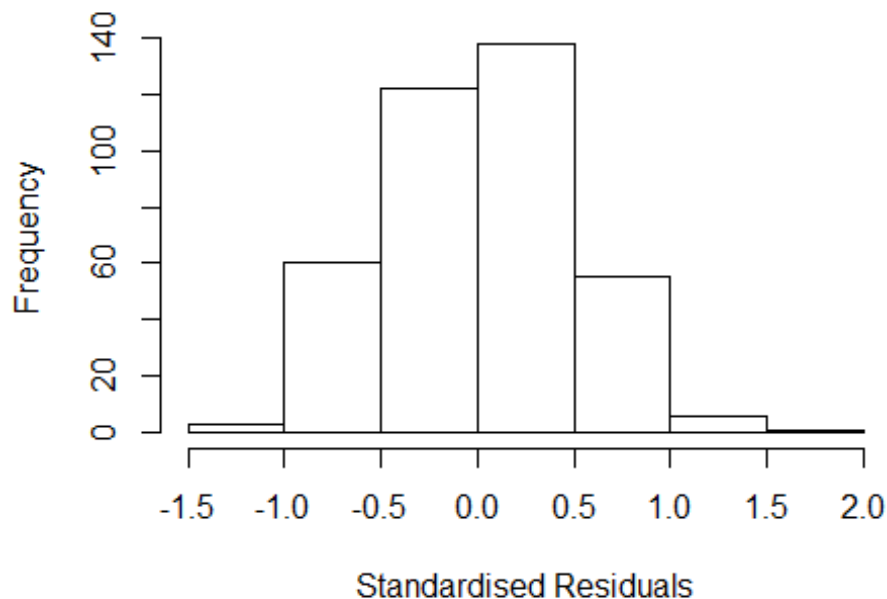
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 190, p-value = 0.4
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
## Year as factors"
##
##          GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.416 1          1.190
## LastAuthorFemale 1.353 1          1.163
## UniqueAuthors    2.231 4          1.106
## Year              2.759 16         1.032
```



## Residuals from first and last author and team size



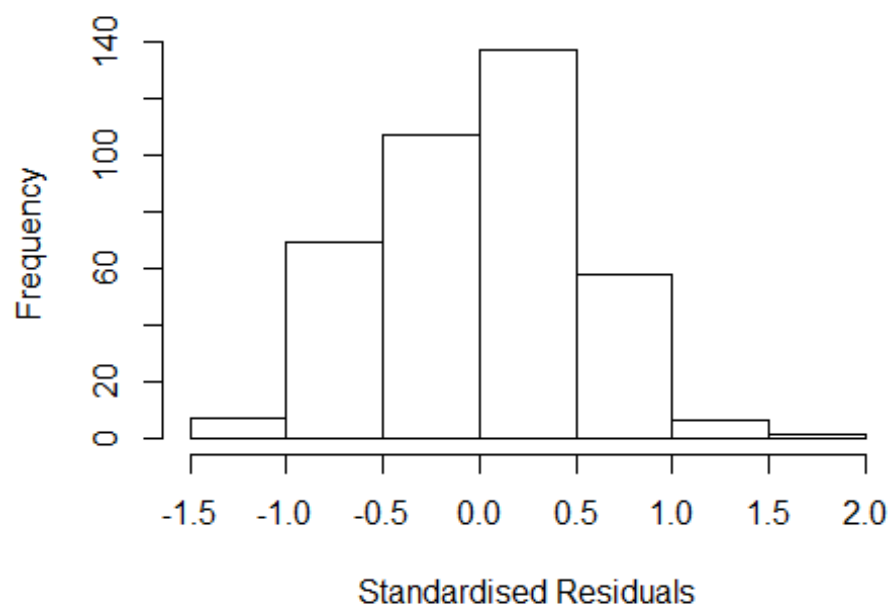
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4474 -0.3849 0.0478 0.3354 1.7096
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.756075 0.169700 4.46 1.1e-05 ***
## FirstAuthorFemale1 0.213549 0.061388 3.48 0.00057 ***
## LastAuthorFemale1 0.036432 0.058505 0.62 0.53387
## UniqueAuthors2 0.308748 0.086335 3.58 0.00040 ***
## UniqueAuthors3 0.477796 0.091722 5.21 3.2e-07 ***
## UniqueAuthors4 0.556090 0.104447 5.32 1.8e-07 ***
## UniqueAuthors5 0.521368 0.094831 5.50 7.3e-08 ***
## Year1997 0.053392 0.203919 0.26 0.79360
## Year1998 0.071581 0.243366 0.29 0.76883
## Year1999 0.000815 0.188748 0.00 0.99656
```

```

## Year2000      -0.245393    0.196764    -1.25    0.21315
## Year2001      -0.201510    0.215779    -0.93    0.35099
## Year2002      -0.407682    0.178852    -2.28    0.02322 *
## Year2003      -0.476874    0.168718    -2.83    0.00497 **
## Year2004      -0.575994    0.204250    -2.82    0.00507 **
## Year2005      -0.431456    0.187138    -2.31    0.02170 *
## Year2006      -0.182908    0.184495    -0.99    0.32215
## Year2007      -0.295816    0.181128    -1.63    0.10330
## Year2008      -0.534087    0.174985    -3.05    0.00244 **
## Year2009      -0.612421    0.184755    -3.31    0.00101 **
## Year2010      -0.275101    0.172806    -1.59    0.11227
## Year2011      -0.619884    0.204120    -3.04    0.00256 **
## Year2012      -0.345999    0.197742    -1.75    0.08101 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.507
## Multiple R-squared:  0.286, Adjusted R-squared:  0.242
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 21 weights are ~= 1. The remaining 364 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.232  0.869  0.948  0.913  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          2.60e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.295 1          1.138
## LastAuthorFemale  1.201 1          1.096
## Year              1.402 16          1.011

```

## Residuals from first and last author



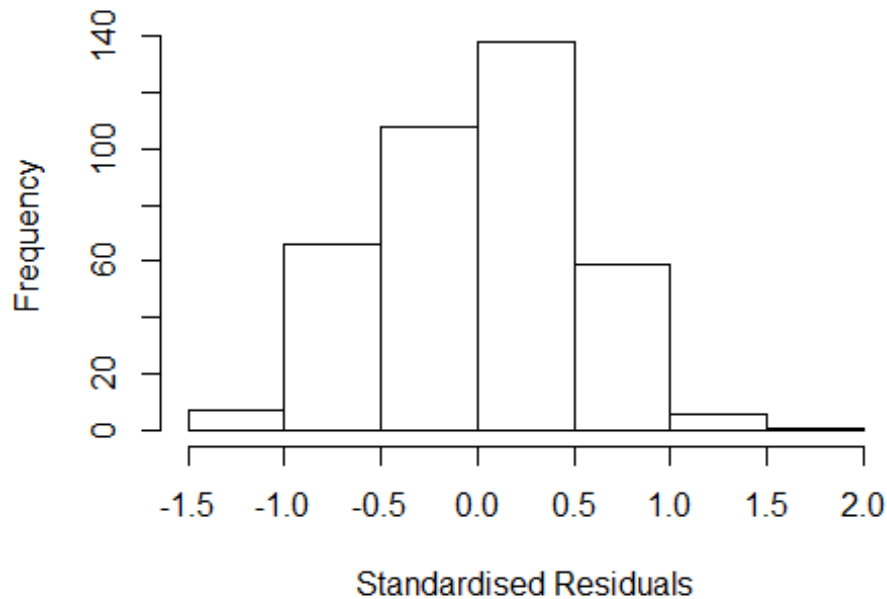
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2581 -0.3979 0.0353 0.3712 1.8452
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9872 0.1510 6.54 2.1e-10 ***
## FirstAuthorFemale1 0.2709 0.0633 4.28 2.4e-05 ***
## LastAuthorFemale1 0.0206 0.0600 0.34 0.7318
## Year1997 0.1725 0.1969 0.88 0.3816
## Year1998 0.1613 0.2295 0.70 0.4824
## Year1999 0.0956 0.1851 0.52 0.6060
## Year2000 -0.1411 0.1918 -0.74 0.4624
## Year2001 0.0273 0.2144 0.13 0.8986
## Year2002 -0.2629 0.1695 -1.55 0.1217
## Year2003 -0.3834 0.1661 -2.31 0.0215 *
## Year2004 -0.4691 0.1959 -2.39 0.0172 *
## Year2005 -0.3197 0.1821 -1.76 0.0800 .
```

```

## Year2006          -0.0333      0.1686   -0.20   0.8436
## Year2007          -0.0804      0.1598   -0.50   0.6150
## Year2008          -0.4179      0.1714   -2.44   0.0152 *
## Year2009          -0.5014      0.1779   -2.82   0.0051 **
## Year2010          -0.1179      0.1659   -0.71   0.4779
## Year2011          -0.5893      0.2000   -2.95   0.0034 **
## Year2012          -0.2137      0.1850   -1.15   0.2489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.536
## Multiple R-squared:  0.199, Adjusted R-squared:  0.16
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 28 weights are ~= 1. The remaining 357 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.212  0.863  0.945  0.912  0.982  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.60e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.225 1      1.107
## Year              1.225 16      1.006

```

## Residuals from first author

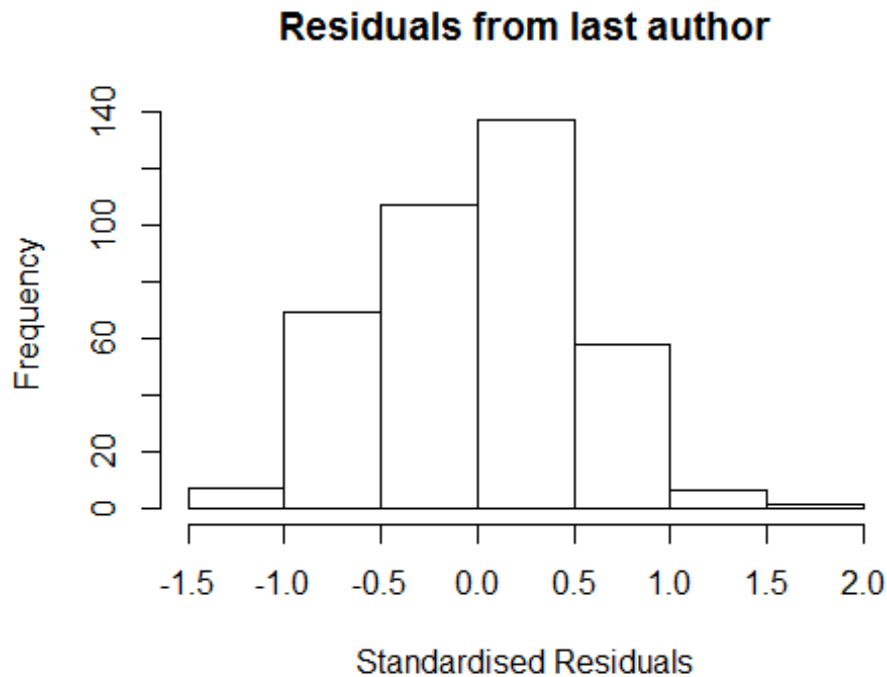


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
## AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.2679 -0.4049 0.0393 0.3802 1.8415
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9913 0.1506 6.58 1.6e-10 ***
## FirstAuthorFemale1 0.2766 0.0617 4.48 1.0e-05 ***
## Year1997 0.1742 0.1976 0.88 0.3786
## Year1998 0.1613 0.2297 0.70 0.4829
## Year1999 0.1014 0.1849 0.55 0.5840
## Year2000 -0.1426 0.1925 -0.74 0.4594
## Year2001 0.0261 0.2152 0.12 0.9037
## Year2002 -0.2622 0.1697 -1.54 0.1232
## Year2003 -0.3817 0.1668 -2.29 0.0227 *
## Year2004 -0.4655 0.1964 -2.37 0.0183 *
## Year2005 -0.3171 0.1829 -1.73 0.0839 .
## Year2006 -0.0343 0.1695 -0.20 0.8396
```

```

## Year2007          -0.0795      0.1607   -0.49   0.6210
## Year2008          -0.4192      0.1721   -2.44   0.0153 *
## Year2009          -0.5018      0.1784   -2.81   0.0052 **
## Year2010          -0.1171      0.1667   -0.70   0.4828
## Year2011          -0.5864      0.2007   -2.92   0.0037 **
## Year2012          -0.2121      0.1853   -1.14   0.2532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.534
## Multiple R-squared:  0.199, Adjusted R-squared:  0.162
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 30 weights are ~= 1. The remaining 355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.210  0.860  0.948  0.911  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.60e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.154 1          1.074
## Year            1.154 16          1.004

```



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3445 -0.4035 0.0433 0.3965 1.6963
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.18547 0.15161 7.82 5.7e-14 ***
## LastAuthorFemale1 0.08899 0.05999 1.48 0.13882
## Year1997 0.07004 0.21091 0.33 0.74000
## Year1998 0.07188 0.21808 0.33 0.74190
## Year1999 0.00781 0.19523 0.04 0.96810
## Year2000 -0.15413 0.21036 -0.73 0.46423
## Year2001 -0.06902 0.21746 -0.32 0.75114
## Year2002 -0.32375 0.18770 -1.72 0.08541 .
## Year2003 -0.52386 0.17012 -3.08 0.00223 **
## Year2004 -0.51756 0.19341 -2.68 0.00778 **
## Year2005 -0.41621 0.18564 -2.24 0.02556 *
## Year2006 -0.11996 0.16841 -0.71 0.47671
```

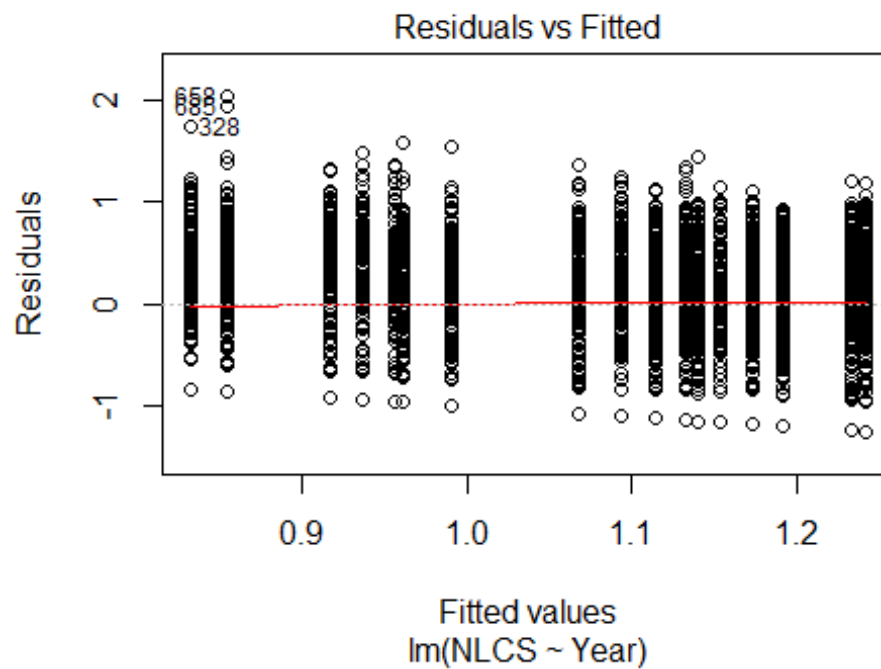
```

## Year2007          -0.10476      0.16632    -0.63   0.52918
## Year2008          -0.44961      0.18017    -2.50   0.01302 *
## Year2009          -0.55081      0.18393    -2.99   0.00293 **
## Year2010          -0.20149      0.17736    -1.14   0.25667
## Year2011          -0.68769      0.20450    -3.36   0.00085 ***
## Year2012          -0.29160      0.19309    -1.51   0.13186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.545
## Multiple R-squared:  0.156, Adjusted R-squared:  0.117
## Convergence in 12 IRWLS iterations
##
## Robustness weights:
## 25 weights are ~= 1. The remaining 360 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.311  0.870   0.944   0.911   0.981   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.60e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 385"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3612"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 289 305 333 293 353 310 331 283 273 291 314 359 380 381 390
## 2011 2012
## 412 438
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 166 199 214 195 189 146 233 200 172 196 236 270 297 302 297
## 2011 2012

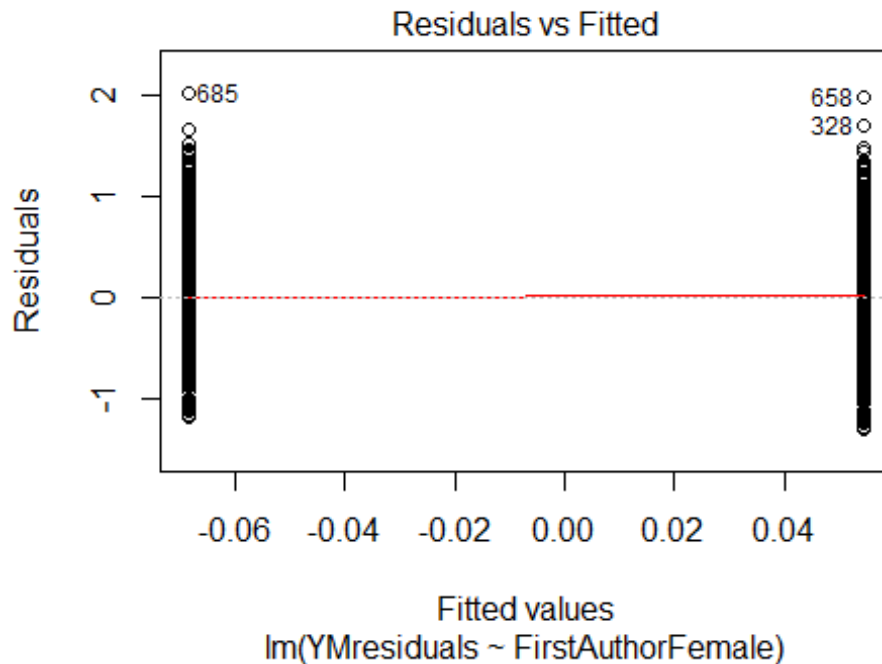
```



```
## 340 354
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 158 187 201 174 168 136 207 170 157 177 215 249 258 277 277
## 2011 2012
## 320 324
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 74, df = 16, p-value = 2e-09
```

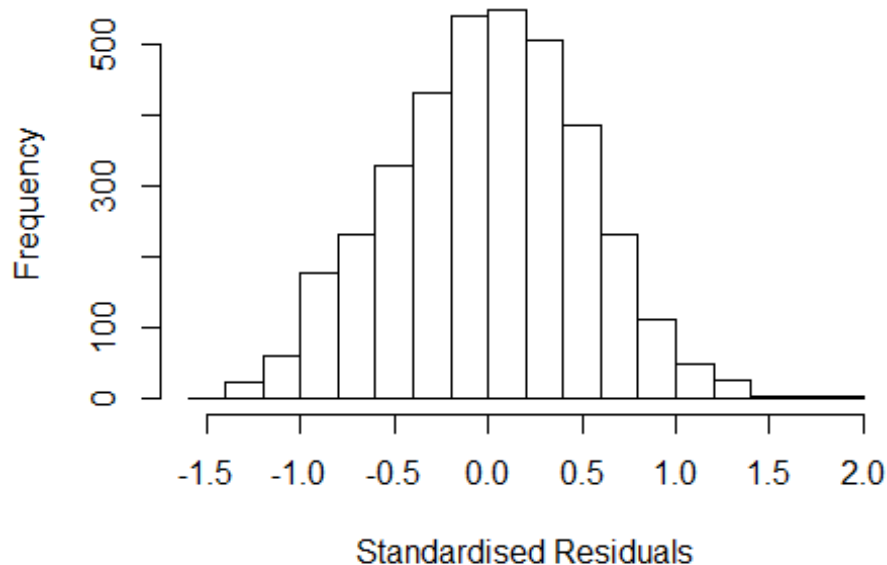


```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 12, df = 1, p-value = 7e-04
```



```
## [1] "Female first author team size 2018 geometric mean: 3.43672829615007"
## [1] "Male first author team size 2018 geometric mean: 3.12833085140042"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 22000, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 3.15015473040331"
## [1] "Male last author team size 2018 geometric mean: 3.28308438821595"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 17000, p-value = 0.3
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.185 1          1.089
## LastAuthorFemale  1.176 1          1.084
## UniqueAuthors    1.240 4          1.027
## Year              1.265 16         1.007
```

## Residuals from first and last author and team size



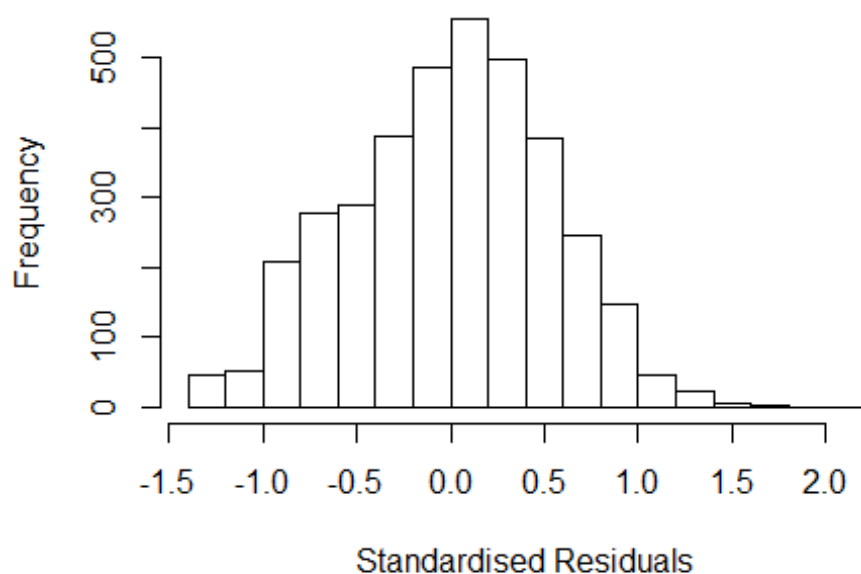
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4318 -0.3541  0.0134  0.3518  1.9634
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.87313    0.05742   15.21 < 2e-16 ***
## FirstAuthorFemale1 -0.09117    0.01937   -4.71 2.6e-06 ***
## LastAuthorFemale1 -0.12484    0.01990   -6.27 3.9e-10 ***
## UniqueAuthors2     0.29591    0.02673   11.07 < 2e-16 ***
## UniqueAuthors3     0.36709    0.02692   13.64 < 2e-16 ***
## UniqueAuthors4     0.40950    0.02985   13.72 < 2e-16 ***
## UniqueAuthors5     0.48057    0.03054   15.74 < 2e-16 ***
## Year1997          -0.06174    0.07189   -0.86  0.391
## Year1998          -0.11540    0.07194   -1.60  0.109
## Year1999          -0.12265    0.07406   -1.66  0.098 .
```

```

## Year2000      -0.06329    0.07247   -0.87    0.383
## Year2001      -0.07169    0.07498   -0.96    0.339
## Year2002       0.03752    0.06753    0.56    0.579
## Year2003      -0.02857    0.06897   -0.41    0.679
## Year2004       0.04347    0.06937    0.63    0.531
## Year2005       0.00174    0.06731    0.03    0.979
## Year2006       0.06007    0.06539    0.92    0.358
## Year2007       0.04252    0.06615    0.64    0.520
## Year2008       0.09428    0.06324    1.49    0.136
## Year2009       0.15494    0.06394    2.42    0.015 *
## Year2010       0.14365    0.06373    2.25    0.024 *
## Year2011       0.02847    0.06524    0.44    0.663
## Year2012       0.07814    0.06398    1.22    0.222
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.519
## Multiple R-squared:  0.176, Adjusted R-squared:  0.171
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 300 weights are ~= 1. The remaining 3355 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.121  0.871  0.950  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00      5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.171 1      1.082
## LastAuthorFemale  1.170 1      1.082
## Year              1.044 16      1.001

```

## Residuals from first and last author



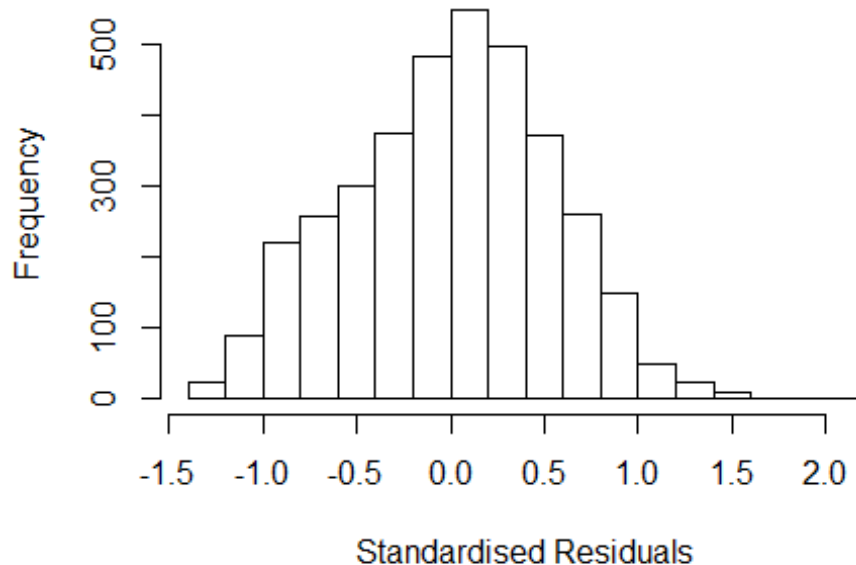
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3434 -0.3784 0.0291 0.3729 2.1083
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0210 0.0611 16.71 < 2e-16 ***
## FirstAuthorFemale1 -0.0755 0.0204 -3.71 0.00021 ***
## LastAuthorFemale1 -0.1618 0.0209 -7.74 1.3e-14 ***
## Year1997 -0.0696 0.0770 -0.90 0.36610
## Year1998 -0.0910 0.0771 -1.18 0.23781
## Year1999 -0.0552 0.0796 -0.69 0.48824
## Year2000 0.0140 0.0762 0.18 0.85444
## Year2001 0.0231 0.0807 0.29 0.77496
## Year2002 0.1423 0.0735 1.94 0.05299 .
## Year2003 0.0639 0.0741 0.86 0.38846
## Year2004 0.2037 0.0736 2.77 0.00564 **
## Year2005 0.1374 0.0711 1.93 0.05328 .
```

```

## Year2006          0.2090      0.0686      3.05  0.00231 **
## Year2007          0.1866      0.0691      2.70  0.00699 **
## Year2008          0.2679      0.0668      4.01  6.2e-05 ***
## Year2009          0.3224      0.0671      4.80  1.6e-06 ***
## Year2010          0.2969      0.0671      4.42  1.0e-05 ***
## Year2011          0.1971      0.0683      2.89  0.00391 **
## Year2012          0.2713      0.0671      4.04  5.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.546
## Multiple R-squared:  0.091, Adjusted R-squared:  0.0865
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 310 weights are ~= 1. The remaining 3345 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.104  0.861  0.948  0.910  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##      trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.023 1      1.012
## Year              1.023 16      1.001

```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.318 -0.382 0.025 0.374 2.054
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.9722 0.0619 15.71 < 2e-16 ***
## FirstAuthorFemale1 -0.1472 0.0194 -7.59 3.9e-14 ***
## Year1997 -0.0607 0.0790 -0.77 0.44244
## Year1998 -0.0780 0.0790 -0.99 0.32352
## Year1999 -0.0385 0.0813 -0.47 0.63577
## Year2000 0.0264 0.0781 0.34 0.73541
## Year2001 0.0291 0.0826 0.35 0.72468
## Year2002 0.1614 0.0749 2.15 0.03133 *
## Year2003 0.0802 0.0755 1.06 0.28836
## Year2004 0.2267 0.0744 3.05 0.00234 **
## Year2005 0.1570 0.0724 2.17 0.03025 *
## Year2006 0.2295 0.0696 3.30 0.00098 ***
```

```

## Year2007          0.2133      0.0703      3.04  0.00242 **
## Year2008          0.2862      0.0680      4.21  2.6e-05 ***
## Year2009          0.3460      0.0681      5.08  4.0e-07 ***
## Year2010          0.3231      0.0681      4.75  2.2e-06 ***
## Year2011          0.2201      0.0690      3.19  0.00144 **
## Year2012          0.2887      0.0682      4.23  2.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.552
## Multiple R-squared:  0.0756, Adjusted R-squared:  0.0713
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 292 weights are ~= 1. The remaining 3363 ones are summarized as
##   Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.136  0.867  0.949  0.910  0.986  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.024 1      1.012
## Year      1.024 16      1.001
##
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField Fields      residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##      control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min      1Q  Median      3Q      Max
## -1.319 -0.375  0.028  0.377  2.090

```



```

##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.9984    0.0610  16.38 < 2e-16 ***
## LastAuthorFemale1 -0.1968    0.0198  -9.94 < 2e-16 ***
## Year1997        -0.0697    0.0774  -0.90  0.3679
## Year1998        -0.0909    0.0772  -1.18  0.2392
## Year1999        -0.0487    0.0801  -0.61  0.5436
## Year2000         0.0180    0.0765   0.23  0.8144
## Year2001         0.0241    0.0815   0.30  0.7678
## Year2002         0.1436    0.0736   1.95  0.0511 .
## Year2003         0.0601    0.0744   0.81  0.4199
## Year2004         0.2024    0.0739   2.74  0.0062 **
## Year2005         0.1397    0.0713   1.96  0.0503 .
## Year2006         0.2152    0.0688   3.13  0.0018 **
## Year2007         0.1894    0.0692   2.74  0.0062 **
## Year2008         0.2718    0.0669   4.06  4.9e-05 ***
## Year2009         0.3210    0.0671   4.78  1.8e-06 ***
## Year2010         0.2979    0.0671   4.44  9.3e-06 ***
## Year2011         0.2044    0.0685   2.98  0.0029 **
## Year2012         0.2781    0.0670   4.15  3.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.55
## Multiple R-squared:  0.087, Adjusted R-squared:  0.0827
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 303 weights are ~= 1. The remaining 3352 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.117  0.866  0.949  0.911  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      2.74e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 3655"
## [1] ""

```

```

## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3613"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##    29    27    20    20    33    25    14    20    25    14    15    20    10    23    14
## 2011 2012
##    11    33
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     1     1     2     0     2     3     6     8    17     6     9    10     7    12     9
## 2011 2012
##     7     28
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##     1     1     2     0     2     3     4     1    16     4     9     8     6    11     6
## 2011 2012
##     5     22
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
## [1] "Female first author team size 2018 geometric mean: 3.46410161513775"
## [1] "Male first author team size 2018 geometric mean: 3.55689330449006"

## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties

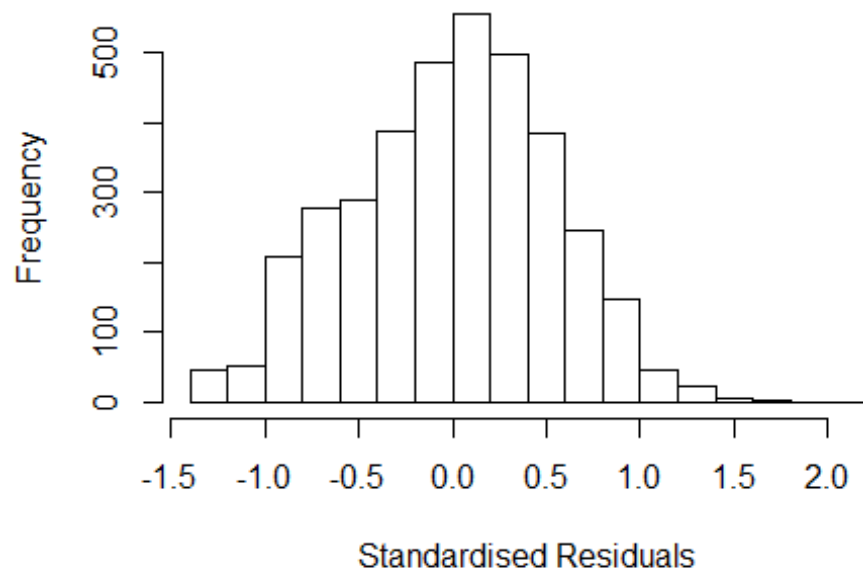
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 3, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: NaN"
## [1] "Male last author team size 2018 geometric mean: 3.51948202893552"
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful

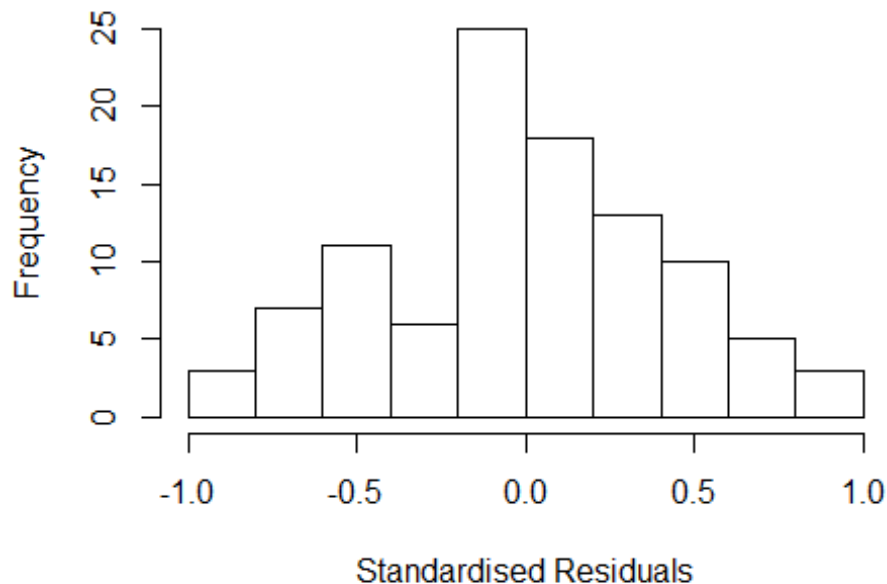
```

## Residuals from last author



##		GVIF	Df	$GVIF^{(1/(2*Df))}$
##	FirstAuthorFemale	NaN	1	NaN
##	LastAuthorFemale	NaN	1	NaN
##	UniqueAuthors	NaN	4	NaN
##	Year	NaN	15	NaN

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -8.87e-01 -3.53e-01 -1.44e-15 2.77e-01 9.83e-01
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8150 0.0000 Inf < 2e-16 ***
## FirstAuthorFemale1 0.2053 0.1088 1.89 0.06286 .
## LastAuthorFemale1 0.0843 0.1166 0.72 0.47174
## UniqueAuthors2 0.1556 0.1786 0.87 0.38632
## UniqueAuthors3 0.3620 0.1755 2.06 0.04247 *
## UniqueAuthors4 0.5885 0.2405 2.45 0.01662 *
## UniqueAuthors5 0.6060 0.2437 2.49 0.01500 *
## Year1997 -0.9730 0.0000 -Inf < 2e-16 ***
## Year1998 -0.7415 0.0980 -7.57 6.0e-11 ***
## Year2000 -0.8553 0.1545 -5.53 3.9e-07 ***
```

```

## Year2001          -1.1046      0.3102    -3.56  0.00063 ***
## Year2002          -1.3359      0.2516    -5.31  9.8e-07 ***
## Year2003          -1.5025      0.2405    -6.25  2.0e-08 ***
## Year2004          -1.3342      0.1843    -7.24  2.6e-10 ***
## Year2005          -1.5251      0.2185    -6.98  8.2e-10 ***
## Year2006          -1.5147      0.1969    -7.69  3.5e-11 ***
## Year2007          -1.0833      0.2924    -3.71  0.00039 ***
## Year2008          -0.7914      0.2766    -2.86  0.00540 **
## Year2009          -1.2283      0.2086    -5.89  9.1e-08 ***
## Year2010          -1.4853      0.3662    -4.06  0.00012 ***
## Year2011          -1.7272      0.1914    -9.02  8.8e-14 ***
## Year2012          -1.3699      0.1864    -7.35  1.6e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.522
## Multiple R-squared:  0.316, Adjusted R-squared:  0.134
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 9 weights are ~= 1. The remaining 92 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.703  0.898  0.954  0.936  0.990  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          9.90e-04          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 46.069 1 6.787
## LastAuthorFemale 1.977 1 1.406
## Year 84.651 15 1.159
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##

```

```

## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1351 -0.3106 -0.0124  0.3613  0.9725
##
## Coefficients:
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept)    1.81e+00   5.53e-08  3.28e+07 < 2e-16 ***
## FirstAuthorFemale1 2.29e-01   1.13e-01  2.02e+00  0.04628 *
## LastAuthorFemale1  8.92e-02   1.15e-01  7.70e-01  0.44176
## Year1997        -9.73e-01   5.25e-08 -1.85e+07 < 2e-16 ***
## Year1998        -6.76e-01   5.73e-02 -1.18e+01 < 2e-16 ***
## Year2000        -7.77e-01   1.86e-01 -4.18e+00  7.2e-05 ***
## Year2001        -8.08e-01   3.16e-01 -2.55e+00  0.01245 *
## Year2002        -1.14e+00   2.25e-01 -5.06e+00  2.5e-06 ***
## Year2003        -9.14e-01   6.05e-08 -1.51e+07 < 2e-16 ***
## Year2004        -1.06e+00   1.61e-01 -6.57e+00  4.1e-09 ***
## Year2005        -1.22e+00   1.54e-01 -7.92e+00  9.3e-12 ***
## Year2006        -1.33e+00   1.38e-01 -9.61e+00  3.9e-15 ***
## Year2007        -6.80e-01   2.42e-01 -2.81e+00  0.00619 **
## Year2008        -5.46e-01   2.50e-01 -2.18e+00  0.03206 *
## Year2009        -9.12e-01   1.30e-01 -7.00e+00  6.2e-10 ***
## Year2010        -1.08e+00   2.76e-01 -3.92e+00  0.00018 ***
## Year2011        -1.40e+00   1.86e-01 -7.52e+00  5.7e-11 ***
## Year2012        -1.03e+00   1.48e-01 -6.99e+00  6.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.542
## Multiple R-squared:  0.221, Adjusted R-squared:  0.0618
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 15 weights are ~= 1. The remaining 86 ones are summarized as
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.640  0.896  0.954  0.928  0.980  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00        5.00e-01      4.69e+00      1.00e-07
##      rel.tol        solve.tol      eps.outlier          eps.x
##      1.00e-07        1.00e-07      9.90e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01        5.00e-01
##      nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500           50          2           1           1000      200
##      trace.lev      mts      compute.rd

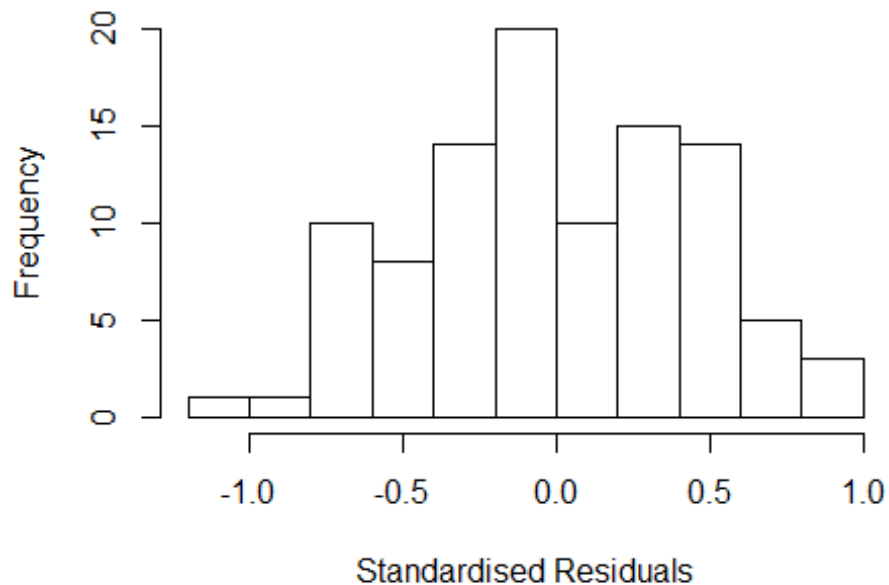
```

```
##           0           1000           0
##           psi           subsampling           cov
##           "bisquare"           "nonsingular"           ".vcov.avar1"
## compute.outlier.stats
##           "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"

## Warning in lf.cov(init, x = x): .vcov.avar1: negative diag(<vcov>) fixed
## up; consider 'cov=".vcov.w."' instead

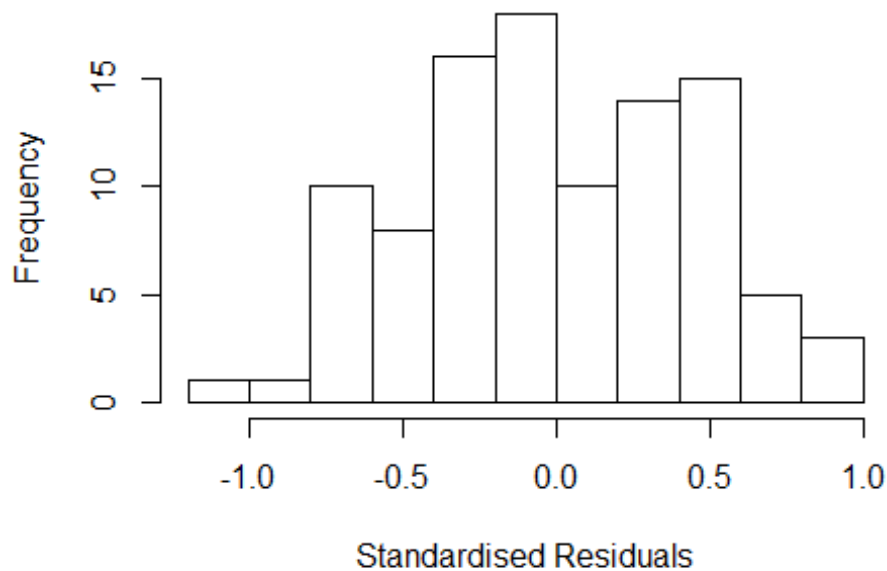
## Warning in lf.cov(init, x = x): diag(.) had 0 or NA entries; non-finite
## result is doubtful
```

### Residuals from first and last author



```
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale NaN 1 NaN
## Year NaN 15 NaN
```

## Residuals from first author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.14999 -0.33117 -0.00459 0.37379 0.96779
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.8150 0.0000 Inf < 2e-16 ***
## FirstAuthorFemale1 0.2448 0.1101 2.22 0.02886 *
## Year1997 -0.9730 0.0000 -Inf < 2e-16 ***
## Year1998 -0.6834 0.0551 -12.39 < 2e-16 ***
## Year2000 -0.7775 0.1858 -4.18 7.0e-05 ***
## Year2001 -0.8075 0.3152 -2.56 0.01220 *
## Year2002 -1.1460 0.2215 -5.17 1.5e-06 ***
## Year2003 -0.9140 0.0000 -Inf < 2e-16 ***
## Year2004 -1.0499 0.1637 -6.41 7.9e-09 ***
## Year2005 -1.2008 0.1412 -8.50 6.0e-13 ***
## Year2006 -1.2909 0.1286 -10.04 4.9e-16 ***
## Year2007 -0.6650 0.2356 -2.82 0.00594 **
```

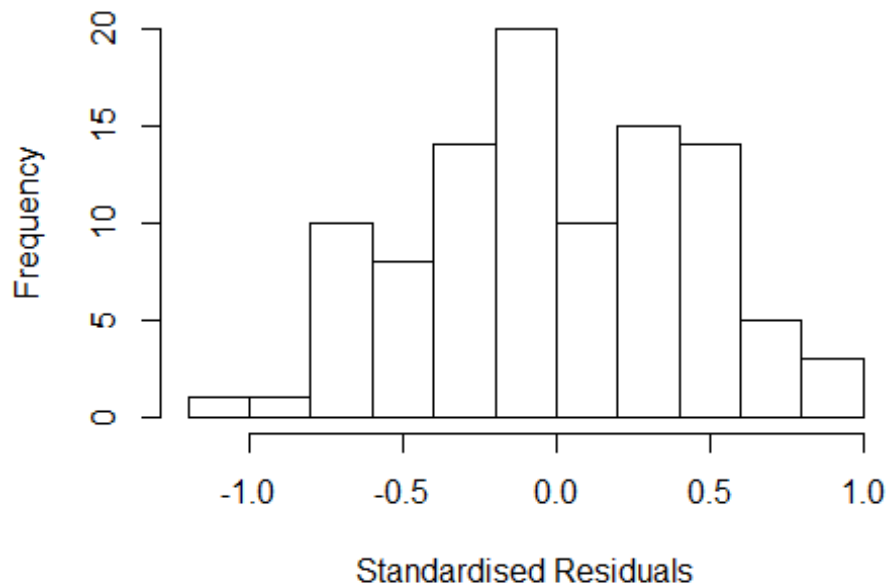


```

## Year2008          -0.5549      0.2470    -2.25   0.02729 *
## Year2009          -0.9099      0.1314    -6.93   8.1e-10 ***
## Year2010          -1.0739      0.2675    -4.01   0.00013 ***
## Year2011          -1.4033      0.1827    -7.68   2.6e-11 ***
## Year2012          -1.0298      0.1461    -7.05   4.7e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.547
## Multiple R-squared:  0.218, Adjusted R-squared:  0.0695
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 16 weights are ~= 1. The remaining 85 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.638  0.901  0.950  0.928  0.981  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.90e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale -9.079 1      NaN
## Year            -9.079 15      NaN

```

## Residuals from last author



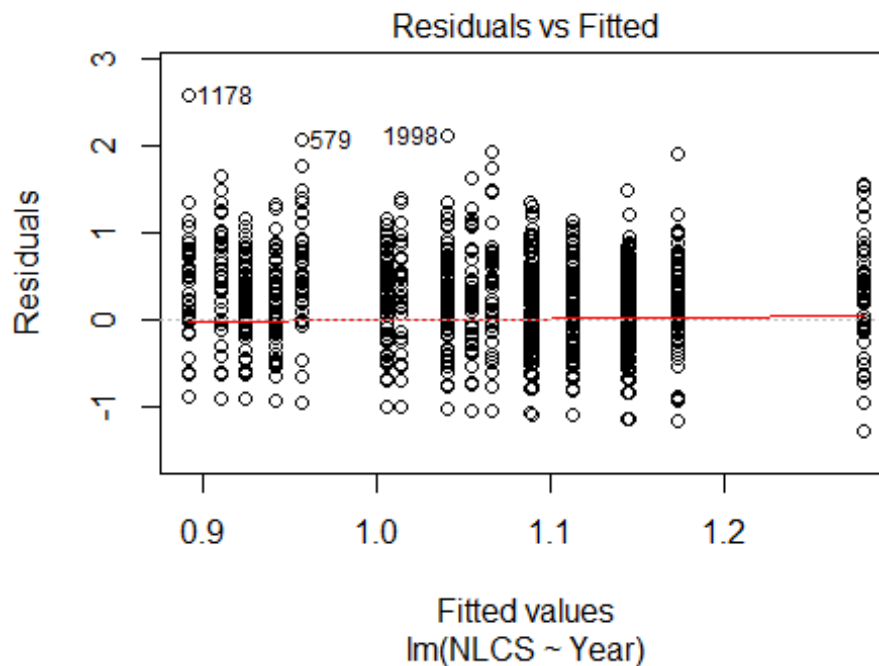
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId  NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26e+00 -3.34e-01  1.11e-16  3.70e-01  1.00e+00
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    1.81e+00   3.91e-08  4.65e+07 < 2e-16 ***
## LastAuthorFemale1 1.63e-01   1.09e-01  1.50e+00  0.13744
## Year1997        -9.73e-01   4.77e-08 -2.04e+07 < 2e-16 ***
## Year1998        -5.61e-01   9.08e-02 -6.18e+00  2.2e-08 ***
## Year2000        -7.77e-01   1.86e-01 -4.17e+00  7.3e-05 ***
## Year2001        -8.08e-01   3.18e-01 -2.54e+00  0.01285 *
## Year2002        -1.01e+00   2.43e-01 -4.15e+00  8.0e-05 ***
## Year2003        -9.14e-01   5.20e-08 -1.76e+07 < 2e-16 ***
## Year2004        -9.46e-01   1.45e-01 -6.52e+00  5.0e-09 ***
## Year2005        -1.20e+00   2.12e-01 -5.65e+00  2.1e-07 ***
## Year2006        -1.19e+00   1.07e-01 -1.11e+01 < 2e-16 ***
## Year2007        -5.56e-01   2.03e-01 -2.73e+00  0.00766 **
```

```

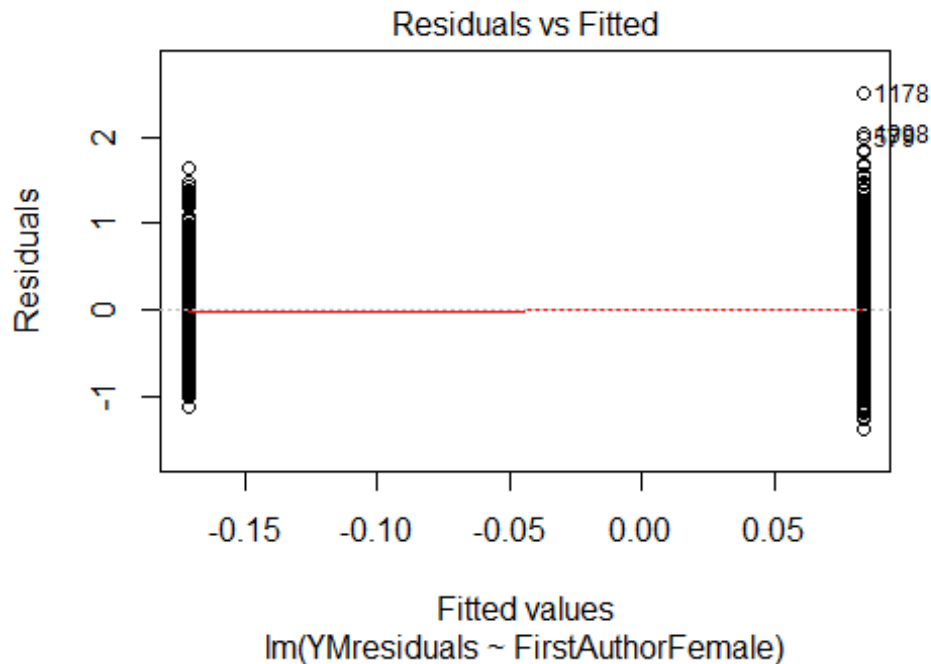
## Year2008          -4.21e-01    2.75e-01 -1.53e+00    0.12893
## Year2009          -8.12e-01    1.05e-01 -7.72e+00    2.2e-11 ***
## Year2010          -9.67e-01    2.53e-01 -3.82e+00    0.00025 ***
## Year2011          -1.30e+00    2.10e-01 -6.18e+00    2.2e-08 ***
## Year2012          -9.49e-01    1.36e-01 -6.99e+00    6.0e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.531
## Multiple R-squared:  0.189, Adjusted R-squared:  0.0347
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 8 weights are ~= 1. The remaining 93 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.553  0.903  0.951  0.927  0.987  0.998
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      9.90e-04      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 101"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3614"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##  157  191  186  190  214  232  201  184  235  203  204  225  210  162  182
## 2011 2012
##  166  135
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   46   60   66   66   56   51   79   81  115  102   89  118  109   88   97
## 2011 2012
##   87   66

```

```
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
##   42   51   59   60   47   45   70   67   84   79   67   89   93   77   87
## 2011 2012
##   82   53
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 74, df = 16, p-value = 2e-09
```



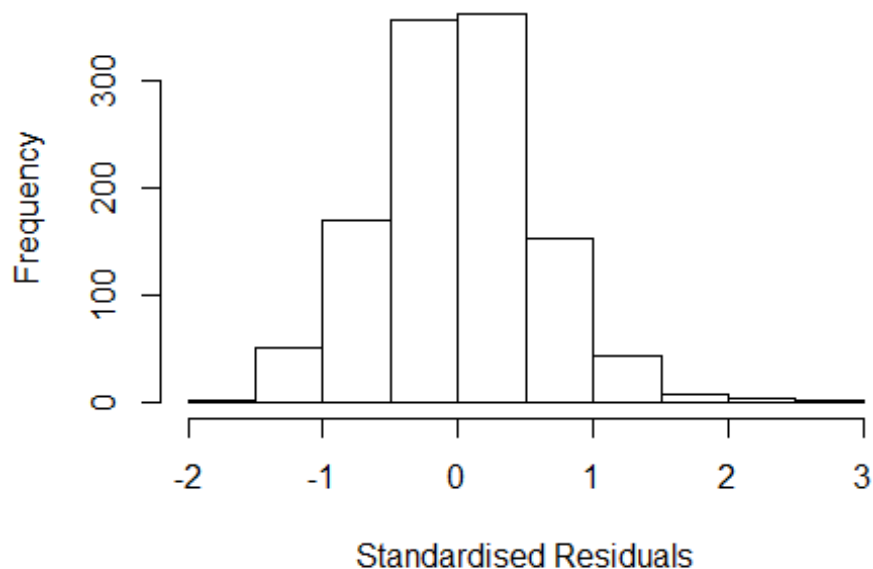
```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 2, df = 1, p-value = 0.2
## [1] "Female first author team size 2018 geometric mean: 4.33765054982491"
## [1] "Male first author team size 2018 geometric mean: 5.84093138706155"
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 400, p-value = 0.02
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 5.3614174677028"
## [1] "Male last author team size 2018 geometric mean: 5.18691132791128"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 570, p-value = 0.6
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF^(1/(2*Df))
FirstAuthorFemale	1.352	1	1.163
LastAuthorFemale	1.613	1	1.270
UniqueAuthors	1.903	4	1.084
Year	1.649	16	1.016

## Residuals from first and last author and team size



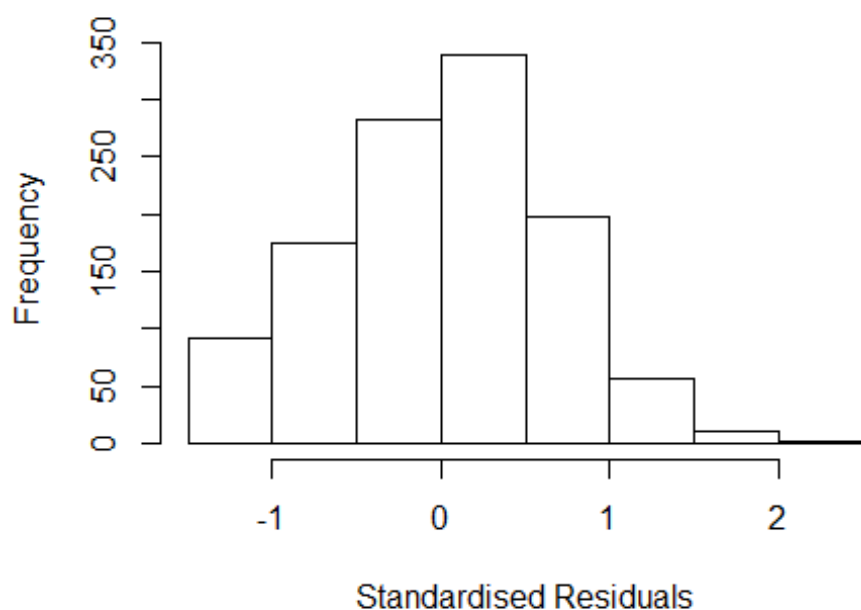
```
## [1] "List of 1 outliers with residuals above 2.5"
##      ScopusId  NLCS Year OneField Fields residuals
## 1178 0035384136 3.474 2001    1704      5    2.538
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
##      Year, data = AllScopusDataOlderFirstLastGendered, control =
## lmrob.control(fast.s.large.n = Inf,
##      k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6119 -0.3614 -0.0091  0.3822  2.5381
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.75445    0.10418   7.24 8.2e-13 ***
## FirstAuthorFemale1 -0.16452    0.04029  -4.08 4.8e-05 ***
## LastAuthorFemale1 -0.20834    0.04688  -4.44 9.7e-06 ***
## UniqueAuthors2     0.40460    0.06612   6.12 1.3e-09 ***
## UniqueAuthors3     0.55419    0.06715   8.25 4.3e-16 ***
## UniqueAuthors4     0.59150    0.06912   8.56 < 2e-16 ***
## UniqueAuthors5     0.89944    0.05882  15.29 < 2e-16 ***
## Year1997          0.06642    0.14601   0.45  0.649
## Year1998         -0.16657    0.12852  -1.30  0.195
## Year1999         -0.10881    0.11961  -0.91  0.363
```

```

## Year2000      -0.21849    0.12543   -1.74    0.082 .
## Year2001      -0.22316    0.13071   -1.71    0.088 .
## Year2002       0.00576    0.11601    0.05    0.960
## Year2003      -0.13333    0.11199   -1.19    0.234
## Year2004      -0.15518    0.12022   -1.29    0.197
## Year2005      -0.03374    0.10742   -0.31    0.753
## Year2006      -0.04199    0.12129   -0.35    0.729
## Year2007      -0.13714    0.11730   -1.17    0.243
## Year2008      -0.00384    0.11457   -0.03    0.973
## Year2009      -0.10607    0.11025   -0.96    0.336
## Year2010      -0.29865    0.11866   -2.52    0.012 *
## Year2011      -0.23774    0.10967   -2.17    0.030 *
## Year2012      -0.29881    0.12578   -2.38    0.018 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.533
## Multiple R-squared:  0.311, Adjusted R-squared:  0.298
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## observation 278 is an outlier with |weight| = 0 ( < 8.7e-05);
## 77 weights are ~= 1. The remaining 1074 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0167 0.8510 0.9510 0.8940 0.9850 0.9990
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.68e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.176 1 1.084
## LastAuthorFemale 1.257 1 1.121
## Year 1.147 16 1.004

```

## Residuals from first and last author



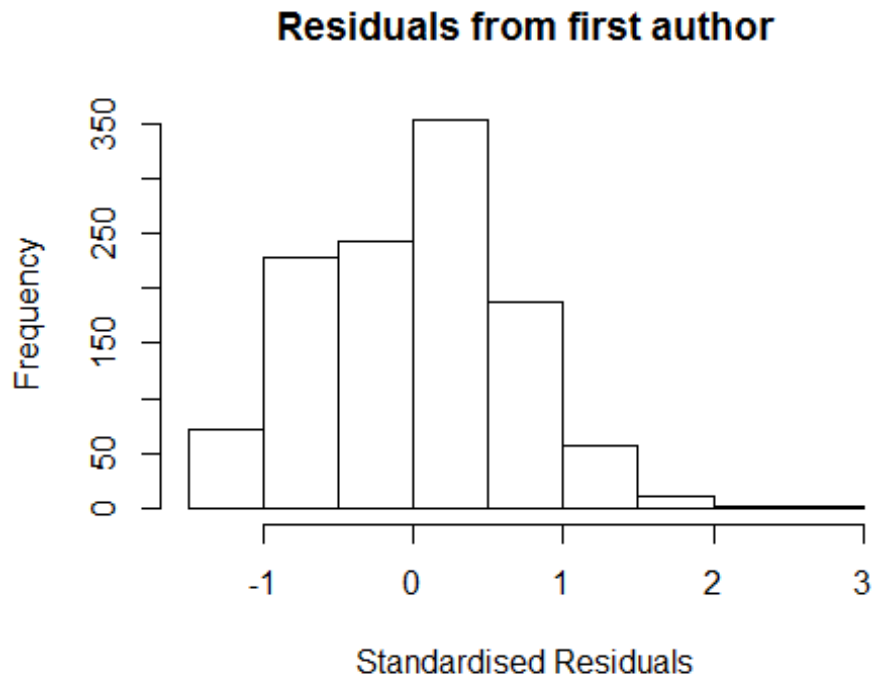
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3748 -0.4497 0.0378 0.4509 2.4547
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.17841 0.10561 11.16 < 2e-16 ***
## FirstAuthorFemale1 -0.17070 0.04519 -3.78 0.00017 ***
## LastAuthorFemale1 -0.40003 0.05708 -7.01 4.1e-12 ***
## Year1997 0.19639 0.16230 1.21 0.22652
## Year1998 -0.16499 0.15428 -1.07 0.28511
## Year1999 0.00377 0.14374 0.03 0.97908
## Year2000 -0.24137 0.16135 -1.50 0.13494
## Year2001 -0.15910 0.15662 -1.02 0.30993
## Year2002 0.08901 0.13341 0.67 0.50477
## Year2003 -0.08453 0.13301 -0.64 0.52522
## Year2004 -0.00693 0.12809 -0.05 0.95684
## Year2005 0.10881 0.11788 0.92 0.35619
```



```

## Year2006          0.08883    0.13098    0.68  0.49782
## Year2007          0.06635    0.12821    0.52  0.60491
## Year2008          0.13198    0.11935    1.11  0.26905
## Year2009          0.12907    0.12360    1.04  0.29657
## Year2010         -0.09813    0.12522   -0.78  0.43342
## Year2011          0.00569    0.12880    0.04  0.96476
## Year2012          0.00542    0.14299    0.04  0.96975
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.641
## Multiple R-squared:  0.121, Adjusted R-squared:  0.107
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 76 weights are ~= 1. The remaining 1076 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.111  0.875  0.947  0.908  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.68e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##   nResample    max.it    best.r.s    k.fast.s      k.max maxit.scale
##      500         50         2         1      1000         200
##   trace.lev      mts    compute.rd
##      0         1000         0
##           psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##           GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.042 1         1.021
## Year              1.042 16         1.001

```

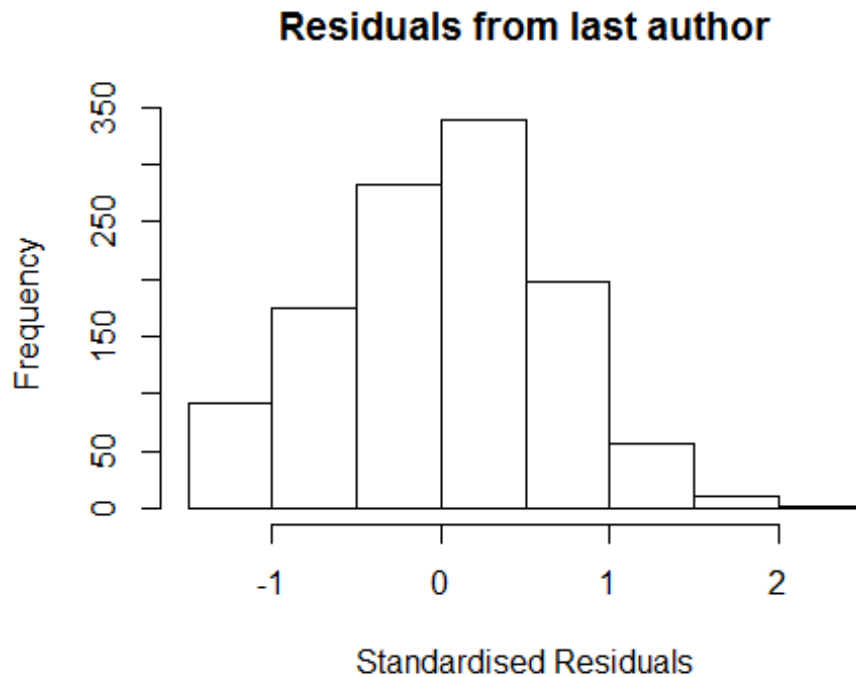


```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3398 -0.5241 0.0555 0.4468 2.5508
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0770 0.1163 9.26 < 2e-16 ***
## FirstAuthorFemale1 -0.2948 0.0456 -6.47 1.5e-10 ***
## Year1997 0.2628 0.1814 1.45 0.15
## Year1998 -0.1809 0.1692 -1.07 0.29
## Year1999 0.0543 0.1592 0.34 0.73
## Year2000 -0.1854 0.1760 -1.05 0.29
## Year2001 -0.1538 0.1754 -0.88 0.38
## Year2002 0.1601 0.1452 1.10 0.27
## Year2003 -0.0277 0.1430 -0.19 0.85
## Year2004 0.0348 0.1384 0.25 0.80
## Year2005 0.2060 0.1284 1.60 0.11
## Year2006 0.1771 0.1387 1.28 0.20
```

```

## Year2007          0.1702      0.1365      1.25      0.21
## Year2008          0.1990      0.1295      1.54      0.12
## Year2009          0.1737      0.1343      1.29      0.20
## Year2010         -0.0432      0.1346     -0.32      0.75
## Year2011          0.0618      0.1413      0.44      0.66
## Year2012          0.1046      0.1498      0.70      0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.665
## Multiple R-squared:  0.0745, Adjusted R-squared:  0.0606
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 84 weights are ~= 1. The remaining 1068 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.109  0.867  0.944  0.909  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.68e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
##  nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
##  trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.093 1      1.045
## Year              1.093 16      1.003

```



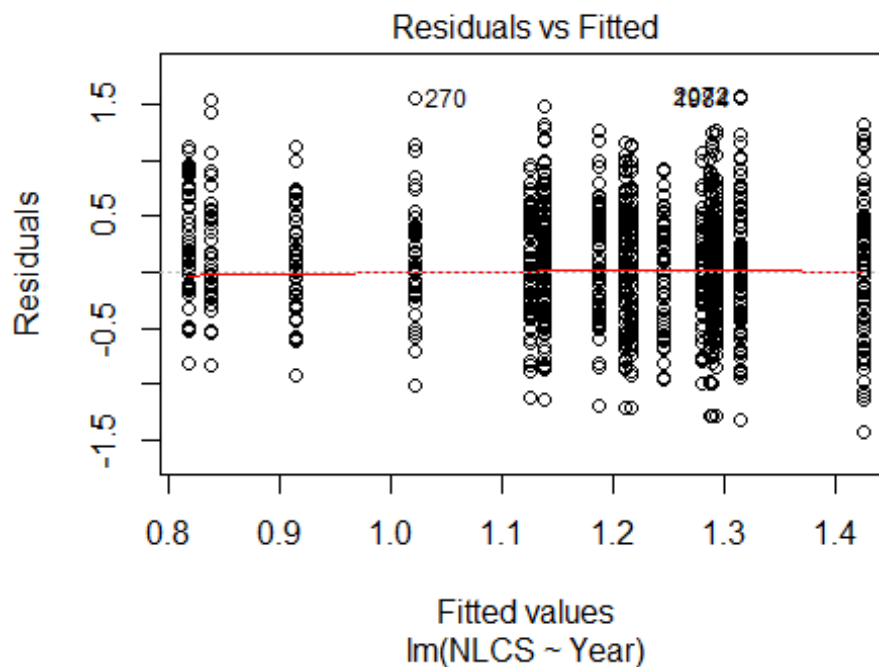
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.3286 -0.4904 0.0376 0.4622 2.5068
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.14877 0.11107 10.34 <2e-16 ***
## LastAuthorFemale1 -0.47892 0.05684 -8.43 <2e-16 ***
## Year1997 0.17983 0.16834 1.07 0.29
## Year1998 -0.17867 0.16164 -1.11 0.27
## Year1999 -0.00743 0.15031 -0.05 0.96
## Year2000 -0.25727 0.16491 -1.56 0.12
## Year2001 -0.18152 0.15866 -1.14 0.25
## Year2002 0.08610 0.13833 0.62 0.53
## Year2003 -0.09052 0.13733 -0.66 0.51
## Year2004 -0.01866 0.13295 -0.14 0.89
## Year2005 0.09199 0.12330 0.75 0.46
## Year2006 0.06521 0.13682 0.48 0.63
```

```

## Year2007      0.05210      0.13251      0.39      0.69
## Year2008      0.11571      0.12434      0.93      0.35
## Year2009      0.11026      0.12819      0.86      0.39
## Year2010     -0.10708      0.12998     -0.82      0.41
## Year2011     -0.00703      0.13341     -0.05      0.96
## Year2012     -0.01864      0.14645     -0.13      0.90
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.643
## Multiple R-squared:  0.11,   Adjusted R-squared:  0.0962
## Convergence in 13 IRWLS iterations
##
## Robustness weights:
## 87 weights are ~= 1. The remaining 1065 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0942 0.8710 0.9470 0.9060 0.9820 0.9990
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.68e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1152"
## [1] ""
## [1] ""
## [1] "#####"
## [1] "Analysis of AJSC 3616"
## [1] "#####"
## [1] "Sample sizes for all years [All, first gendered, first & last
gendered] [check that these decrease]"
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 143 151 103 65 79 74 83 88 107 100 121 129 103 118 104
## 2011 2012
## 124 132
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 58 61 83 54 60 48 69 70 80 75 96 110 90 100 92
## 2011 2012

```

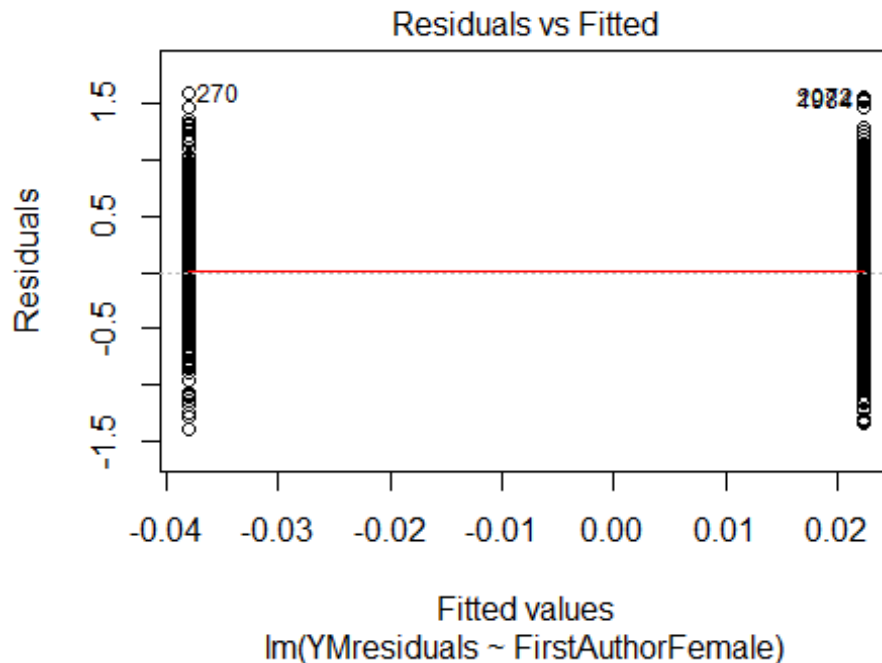
```
## 107 112
##
## 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
## 56 58 74 51 54 44 61 58 74 68 81 99 77 94 85
## 2011 2012
## 101 107
## [1] "Heteroscedasticity checks, confirming that there are problems with
these"
##
## Bartlett test of homogeneity of variances
##
## data: NLCS by Year
## Bartlett's K-squared = 24, df = 16, p-value = 0.08
```



```
##
## Bartlett test of homogeneity of variances
##
## data: YMresiduals by FirstAuthorFemale
## Bartlett's K-squared = 4.3, df = 1, p-value = 0.04
##
## [1] "Female first author team size 2018 geometric mean: 3.05967270445748"
## [1] "Male first author team size 2018 geometric mean: 2.77068221128006"
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 890, p-value = 0.2
```

```
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Female last author team size 2018 geometric mean: 2.75724162998686"
## [1] "Male last author team size 2018 geometric mean: 3.21700857639671"

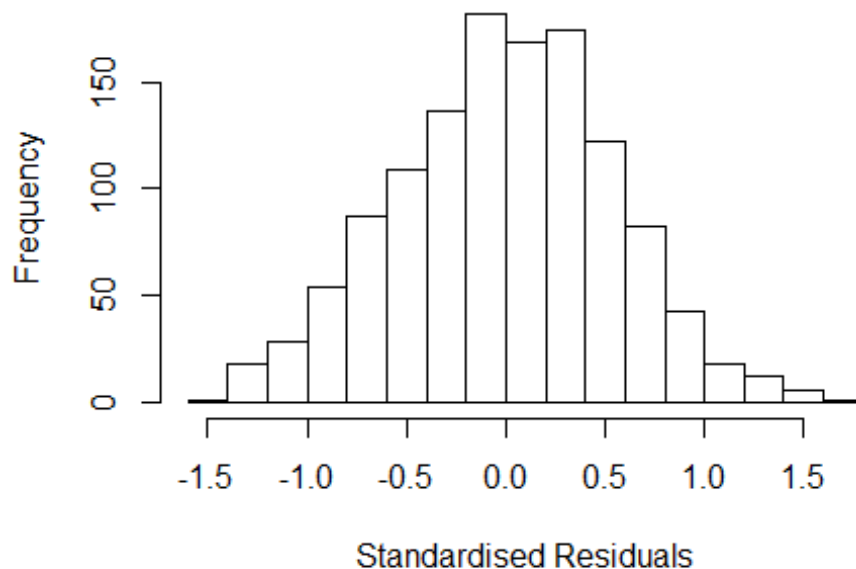
## Warning in wilcox.test.default(FemaleTeamSizes2018, MaleTeamSizes2018,
## alternative = "two.sided"): cannot compute exact p-value with ties
```



```
##
## Wilcoxon rank sum test with continuity correction
##
## data: FemaleTeamSizes2018 and MaleTeamSizes2018
## W = 730, p-value = 0.2
## alternative hypothesis: true location shift is not equal to 0
##
## [1] "Regression 1: First author gender, last author gender, team size,
Year as factors"
##
```

	GVIF	Df	GVIF <sup>1/(2*Df)</sup>
FirstAuthorFemale	1.172	1	1.083
LastAuthorFemale	1.211	1	1.100
UniqueAuthors	1.449	4	1.047
Year	1.556	16	1.014

## Residuals from first and last author and team size



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale +
## UniqueAuthors +
## Year, data = AllScopusDataOlderFirstLastGendered, control =
lmrob.control(fast.s.large.n = Inf,
## k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4193 -0.3736 0.0116 0.3589 1.6022
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.66977 0.08800 7.61 5.4e-14 ***
## FirstAuthorFemale1 0.00483 0.03686 0.13 0.89572
## LastAuthorFemale1 0.02441 0.03496 0.70 0.48529
## UniqueAuthors2 0.21163 0.04556 4.64 3.8e-06 ***
## UniqueAuthors3 0.23351 0.04730 4.94 9.0e-07 ***
## UniqueAuthors4 0.26556 0.05897 4.50 7.3e-06 ***
## UniqueAuthors5 0.16329 0.06748 2.42 0.01566 *
## Year1997 0.16545 0.12176 1.36 0.17444
## Year1998 -0.08031 0.10284 -0.78 0.43499
## Year1999 0.29301 0.11714 2.50 0.01250 *
```

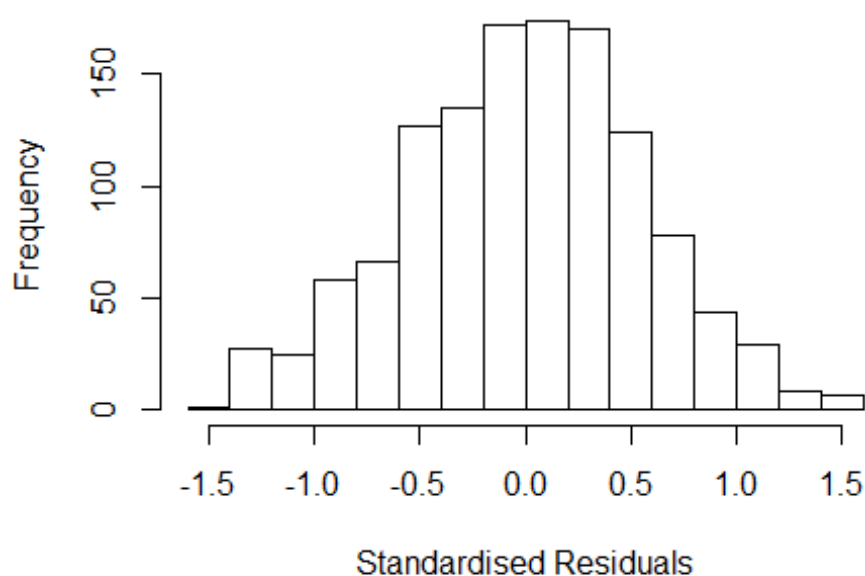


```

## Year2000      0.37338      0.10641      3.51  0.00047 ***
## Year2001      0.00735      0.11245      0.07  0.94791
## Year2002      0.40157      0.09810      4.09  4.5e-05 ***
## Year2003      0.47918      0.11989      4.00  6.8e-05 ***
## Year2004      0.56107      0.10740      5.22  2.1e-07 ***
## Year2005      0.42759      0.10990      3.89  0.00011 ***
## Year2006      0.26832      0.10216      2.63  0.00873 **
## Year2007      0.43357      0.09411      4.61  4.5e-06 ***
## Year2008      0.30712      0.09913      3.10  0.00199 **
## Year2009      0.38193      0.09903      3.86  0.00012 ***
## Year2010      0.34159      0.10638      3.21  0.00136 **
## Year2011      0.37655      0.09743      3.87  0.00012 ***
## Year2012      0.45347      0.10325      4.39  1.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.553
## Multiple R-squared:  0.114, Adjusted R-squared:  0.0978
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 106 weights are ~= 1. The remaining 1136 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.381  0.872  0.951  0.909  0.985  0.999
## Algorithmic parameters:
##      tuning.chi          bb          tuning.psi          refine.tol
##      1.55e+00          5.00e-01          4.69e+00          1.00e-07
##      rel.tol          solve.tol          eps.outlier          eps.x
##      1.00e-07          1.00e-07          8.05e-05          1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 2: First author gender, Last author gender, Year as
factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.157 1          1.076
## LastAuthorFemale  1.174 1          1.084
## Year              1.117 16          1.003

```

## Residuals from first and last author



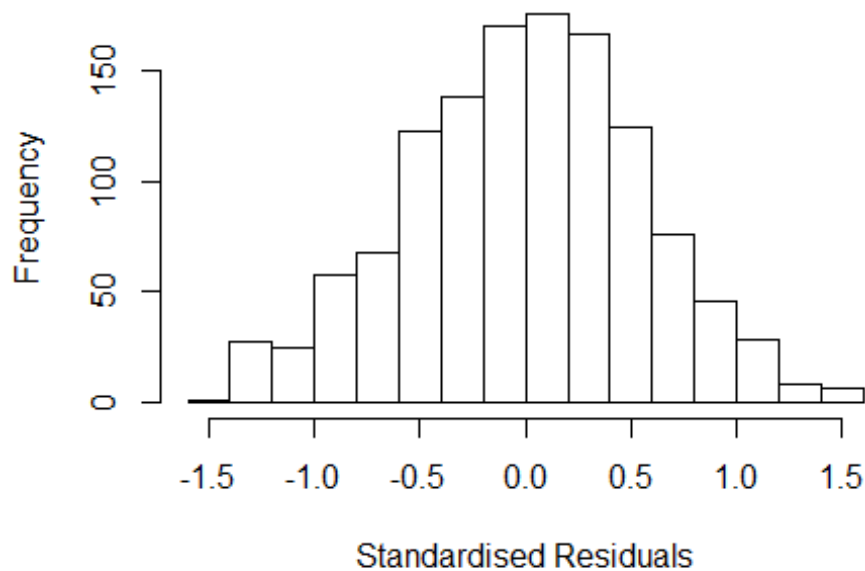
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS      Year      OneField  Fields    residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
##       control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4031 -0.3869  0.0139  0.3717  1.5785
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.8036    0.0858   9.37 < 2e-16 ***
## FirstAuthorFemale1 0.0219    0.0374   0.59  0.55746
## LastAuthorFemale1 0.0110    0.0352   0.31  0.75434
## Year1997         0.1799    0.1194   1.51  0.13232
## Year1998        -0.0694    0.1043  -0.67  0.50585
## Year1999         0.3044    0.1186   2.57  0.01041 *
## Year2000         0.4040    0.1081   3.74  0.00019 ***
## Year2001         0.0522    0.1098   0.48  0.63478
## Year2002         0.4550    0.0984   4.62  4.2e-06 ***
## Year2003         0.5014    0.1242   4.04  5.8e-05 ***
## Year2004         0.5995    0.1062   5.64  2.1e-08 ***
## Year2005         0.4794    0.1105   4.34  1.5e-05 ***
```

```

## Year2006          0.2919      0.1016      2.87  0.00415 **
## Year2007          0.4659      0.0936      4.98  7.3e-07 ***
## Year2008          0.3504      0.0993      3.53  0.00044 ***
## Year2009          0.4265      0.0998      4.27  2.1e-05 ***
## Year2010          0.3673      0.1081      3.40  0.00070 ***
## Year2011          0.4112      0.0975      4.22  2.6e-05 ***
## Year2012          0.5019      0.1019      4.93  9.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.558
## Multiple R-squared:  0.0891, Adjusted R-squared:  0.0757
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 119 weights are ~= 1. The remaining 1123 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.403  0.870   0.948   0.907   0.984   0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.05e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 3: First author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## FirstAuthorFemale 1.052 1      1.026
## Year              1.052 16      1.002

```

## Residuals from first author



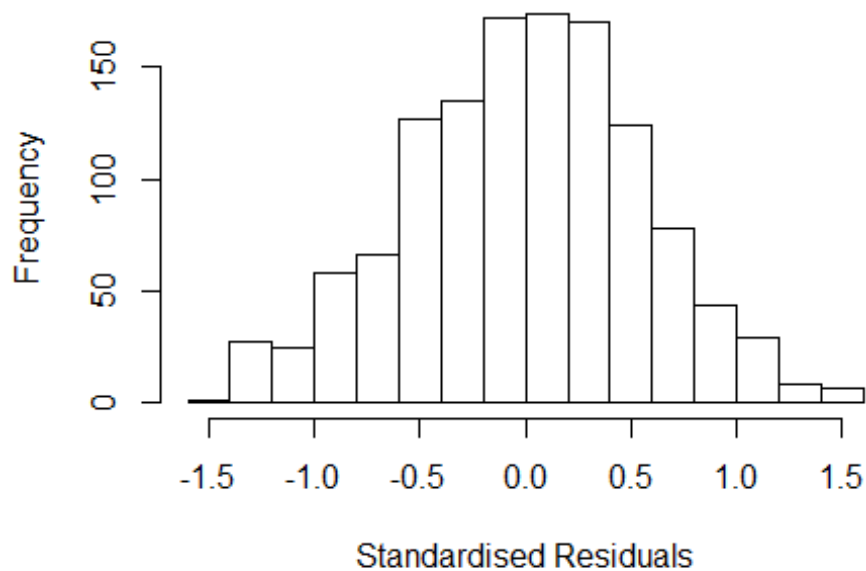
```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ FirstAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4070 -0.3845 0.0106 0.3712 1.5859
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8075 0.0845 9.56 < 2e-16 ***
## FirstAuthorFemale1 0.0258 0.0357 0.72 0.46960
## Year1997 0.1796 0.1193 1.50 0.13268
## Year1998 -0.0688 0.1042 -0.66 0.50940
## Year1999 0.3050 0.1187 2.57 0.01031 *
## Year2000 0.4033 0.1079 3.74 0.00019 ***
## Year2001 0.0527 0.1097 0.48 0.63133
## Year2002 0.4548 0.0984 4.62 4.3e-06 ***
## Year2003 0.5020 0.1243 4.04 5.7e-05 ***
## Year2004 0.5995 0.1062 5.65 2.0e-08 ***
## Year2005 0.4779 0.1103 4.33 1.6e-05 ***
## Year2006 0.2922 0.1017 2.87 0.00414 **
```

```

## Year2007          0.4645      0.0933      4.98  7.3e-07 ***
## Year2008          0.3492      0.0992      3.52  0.00044 ***
## Year2009          0.4257      0.0996      4.28  2.1e-05 ***
## Year2010          0.3675      0.1081      3.40  0.00070 ***
## Year2011          0.4120      0.0975      4.23  2.6e-05 ***
## Year2012          0.5013      0.1018      4.93  9.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.557
## Multiple R-squared:  0.0891, Adjusted R-squared:  0.0764
## Convergence in 11 IRWLS iterations
##
## Robustness weights:
## 122 weights are ~= 1. The remaining 1120 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.398  0.869  0.947  0.906  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol      solve.tol      eps.outlier      eps.x
##      1.00e-07      1.00e-07      8.05e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01      5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000      200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi      subsampling      cov
##      "bisquare"      "nonsingular"      ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Regression 4: Last author gender, Year as factors"
##      GVIF Df GVIF^(1/(2*Df))
## LastAuthorFemale 1.066 1      1.033
## Year              1.066 16      1.002

```

## Residuals from last author



```
## [1] "List of 0 outliers with residuals above 2.5"
## [1] ScopusId NLCS Year OneField Fields residuals
## <0 rows> (or 0-length row.names)
##
## Call:
## lmrob(formula = NLCS ~ LastAuthorFemale + Year, data =
AllScopusDataOlderFirstLastGendered,
## control = lmrob.control(fast.s.large.n = Inf, k.max = 1000))
## \--> method = "MM"
## Residuals:
## Min 1Q Median 3Q Max
## -1.4104 -0.3890 0.0095 0.3674 1.5688
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8148 0.0830 9.82 < 2e-16 ***
## LastAuthorFemale1 0.0182 0.0335 0.54 0.58654
## Year1997 0.1787 0.1190 1.50 0.13369
## Year1998 -0.0685 0.1039 -0.66 0.50999
## Year1999 0.3031 0.1186 2.56 0.01071 *
## Year2000 0.4036 0.1079 3.74 0.00019 ***
## Year2001 0.0505 0.1094 0.46 0.64417
## Year2002 0.4547 0.0982 4.63 4.0e-06 ***
## Year2003 0.4965 0.1235 4.02 6.2e-05 ***
## Year2004 0.5956 0.1057 5.63 2.2e-08 ***
## Year2005 0.4789 0.1106 4.33 1.6e-05 ***
## Year2006 0.2918 0.1013 2.88 0.00402 **
```

```

## Year2007          0.4652      0.0933      4.98  7.1e-07 ***
## Year2008          0.3495      0.0989      3.53  0.00043 ***
## Year2009          0.4263      0.0994      4.29  2.0e-05 ***
## Year2010          0.3633      0.1077      3.37  0.00077 ***
## Year2011          0.4092      0.0972      4.21  2.8e-05 ***
## Year2012          0.5004      0.1016      4.93  9.5e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Robust residual standard error: 0.559
## Multiple R-squared:  0.0888, Adjusted R-squared:  0.0762
## Convergence in 10 IRWLS iterations
##
## Robustness weights:
## 127 weights are ~= 1. The remaining 1115 ones are summarized as
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.411  0.868  0.947  0.906  0.983  0.999
## Algorithmic parameters:
##      tuning.chi          bb      tuning.psi      refine.tol
##      1.55e+00          5.00e-01      4.69e+00      1.00e-07
##      rel.tol          solve.tol      eps.outlier      eps.x
##      1.00e-07          1.00e-07      8.05e-05      1.82e-12
## warn.limit.reject warn.limit.meanrw
##      5.00e-01          5.00e-01
## nResample      max.it      best.r.s      k.fast.s      k.max maxit.scale
##      500          50          2          1          1000          200
## trace.lev      mts      compute.rd
##      0          1000          0
##      psi          subsampling          cov
##      "bisquare"          "nonsingular"          ".vcov.avar1"
## compute.outlier.stats
##      "SM"
## seed : int(0)
## [1] "Sample size for the above analysis: 1242"

for (iRow in 1:FieldCount){
  tryCatch({
    for (iCol in 1:10) {
      if (RegP[iRow,iCol]<0.001) {
        RegStar[iRow,iCol] <- "***"
      } else if (RegP[iRow,iCol]<0.01) {
        RegStar[iRow,iCol] <- "**"
      } else if (RegP[iRow,iCol]<0.05) {
        RegStar[iRow,iCol] <- "*"
      }
    }
  }, error = function(e) return("failed sample size"))
}
print("Stats just for gendered aticles from 1 country 1996 - 2018")

```

```
## [1] "Stats just for gendered aticles from 1 country 1996 - 2018"
```

```
print(BasicStats)
```

##	1Ctry	Gend	1st nth	Fem1 96	Fem1 18	Chg1	FemN 96	FemN 18	ChgN
## 1000	8920		64.283	14.857	36.942	22.0844	9.714	23.883	14.16888
## 1100	16343		65.771	23.348	45.234	21.8865	18.943	32.829	13.88632
## 1101	2228		55.610	25.581	33.929	8.3472	13.953	16.071	2.11794
## 1102	7300		33.986	30.000	41.727	11.7266	12.500	26.619	14.11871
## 1103	13687		42.800	24.878	47.099	22.2209	21.463	33.106	11.64239
## 1104	10722		43.070	22.905	39.035	16.1301	21.229	23.684	2.45516
## 1105	23378		57.781	22.042	39.379	17.3373	19.258	24.020	4.76207
## 1106	8246		47.053	38.693	47.208	8.5147	28.643	39.086	10.44308
## 1107	3173		44.248	11.364	32.955	21.5909	9.091	25.000	15.90909
## 1108	1643		28.911	22.449	38.462	16.0126	6.122	7.692	1.56986
## 1109	4243		51.261	30.097	38.095	7.9982	11.650	27.381	15.73047
## 1110	10266		49.971	30.363	41.791	11.4280	18.482	26.866	8.38382
## 1111	4042		34.834	21.569	26.761	5.1919	11.765	12.676	0.91135
## 1200	2915		66.518	26.667	50.450	23.7838	33.333	49.550	16.21622
## 1201	17185		74.501	33.430	51.936	18.5057	33.721	48.331	14.61018
## 1202	23586		83.613	21.448	37.215	15.7671	22.520	36.821	14.30129
## 1203	13505		79.911	39.373	53.861	14.4886	37.282	52.079	14.79698
## 1204	5276		73.863	28.750	38.865	10.1146	35.000	29.694	-5.30568
## 1205	1555		74.469	17.073	30.769	13.6961	17.073	34.615	17.54221
## 1206	829		78.528	0.000	48.101	48.1013	0.000	46.835	46.83544
## 1207	6503		81.393	25.758	36.422	10.6641	23.485	35.463	11.97841
## 1208	11496		86.500	33.000	48.777	15.7768	33.000	48.777	15.77676
## 1209	444		85.360	0.000	41.176	41.1765	0.000	32.353	32.35294
## 1210	2069		84.050	23.684	35.185	11.5010	23.684	34.259	10.57505
## 1211	13404		84.766	15.487	30.935	15.4485	14.159	30.456	16.29634
## 1212	6482		86.424	27.885	29.040	1.1558	24.038	29.040	5.00194
## 1213	8212		86.617	35.870	52.878	17.0081	35.870	55.396	19.52612
## 1300	21512		65.652	20.621	43.278	22.6563	13.842	31.403	17.56153
## 1301	174		78.736	NaN	33.333	NaN	NaN	30.000	NaN
## 1302	2224		61.061	35.000	55.932	20.9322	25.000	37.288	12.28814
## 1303	41161		63.147	26.808	36.334	9.5260	12.637	18.985	6.34831
## 1304	11089		52.259	23.984	34.286	10.3020	13.008	20.408	7.40003
## 1305	11205		50.245	30.000	34.752	4.7518	12.000	21.986	9.98582
## 1306	12795		43.173	41.111	49.254	8.1426	21.111	27.985	6.87396
## 1307	22263		60.643	34.265	46.437	12.1726	16.765	24.324	7.55962
## 1308	8169		55.454	24.859	46.032	21.1730	18.644	21.429	2.78450
## 1309	5917		58.408	50.360	54.795	4.4348	28.777	37.671	8.89425
## 1310	8552		37.208	29.927	46.429	16.5016	15.328	30.714	15.38582
## 1311	22521		55.046	35.447	43.617	8.1699	20.650	30.213	9.56236
## 1312	34230		61.431	31.480	45.294	13.8145	14.300	25.588	11.28833
## 1313	7526		63.061	23.669	46.696	23.0274	14.793	23.348	8.55512
## 1314	13782		52.242	31.118	42.643	11.5248	12.387	22.823	10.43612
## 1315	4829		66.618	27.815	38.043	10.2289	15.232	27.174	11.94212
## 1400	7643		76.344	15.385	35.870	20.4849	18.376	33.333	14.95726
## 1401	3651		75.431	30.357	35.772	5.4152	31.250	30.894	-0.35569



##	1402	4123	76.134	18.919	32.203	13.2845	16.216	26.554	10.33746
##	1403	9079	74.832	22.000	30.412	8.4124	23.333	29.639	6.30584
##	1404	2554	52.937	13.699	31.034	17.3359	15.068	24.138	9.06944
##	1405	8047	76.563	16.183	37.582	21.3991	17.842	32.026	14.18382
##	1406	5568	70.636	27.642	36.559	8.9169	27.642	38.172	10.52977
##	1407	7199	83.317	25.592	43.612	18.0199	26.540	40.969	14.42888
##	1408	15121	72.006	22.345	36.796	14.4506	21.903	32.218	10.31565
##	1409	4095	79.292	35.165	39.070	3.9049	30.769	30.233	-0.53667
##	1410	1453	76.256	18.519	33.333	14.8148	25.926	33.333	7.40741
##	1500	13346	33.838	15.672	28.253	12.5811	11.194	17.472	6.27809
##	1501	554	37.906	0.000	32.143	32.1429	0.000	14.286	14.28571
##	1502	6231	49.141	22.078	28.788	6.7100	14.286	18.182	3.89610
##	1503	10530	65.698	11.301	28.117	16.8153	7.534	14.058	6.52411
##	1504	223	46.637	50.000	0.000	-50.0000	50.000	50.000	0.00000
##	1505	2902	55.031	17.073	25.333	8.2602	12.195	14.667	2.47154
##	1506	611	33.879	0.000	40.000	40.0000	0.000	40.000	40.00000
##	1507	1939	30.273	13.333	21.795	8.4615	6.667	10.256	3.58974
##	1508	1810	50.166	10.000	22.472	12.4719	3.333	8.989	5.65543
##	1600	34096	55.282	14.319	27.953	13.6342	12.922	17.310	4.38794
##	1601	1092	42.399	14.286	51.020	36.7347	7.143	20.408	13.26531
##	1602	7973	59.112	19.028	36.250	17.2217	12.955	15.625	2.66953
##	1603	4364	49.358	10.000	36.364	26.3636	4.286	19.697	15.41126
##	1604	8088	67.446	14.844	31.844	16.9998	8.203	21.229	13.02593
##	1605	19386	65.944	15.721	32.595	16.8737	10.049	11.709	1.66024
##	1606	21829	58.111	14.136	25.474	11.3376	7.853	14.526	6.67291
##	1607	6201	55.459	20.134	38.889	18.7547	6.040	18.056	12.01529
##	1700	4268	57.779	14.286	17.391	3.1056	12.088	15.528	3.44004
##	1701	577	62.392	NaN	40.741	NaN	NaN	22.222	NaN
##	1702	6679	58.901	17.213	26.190	8.9774	16.393	19.048	2.65418
##	1703	4745	55.279	9.091	21.429	12.3377	6.818	24.107	17.28896
##	1704	2189	51.622	9.091	20.690	11.5987	9.091	17.241	8.15047
##	1705	5854	51.913	8.861	11.579	2.7182	8.861	12.632	3.77082
##	1706	18199	49.096	14.904	18.612	3.7081	19.231	20.505	1.27396
##	1707	2758	46.991	4.545	17.460	12.9149	2.273	19.048	16.77489
##	1708	3164	51.928	14.754	20.000	5.2459	11.475	10.909	-0.56632
##	1709	3211	65.089	26.531	38.889	12.3583	28.571	30.159	1.58730
##	1710	6682	64.517	26.699	23.404	-3.2948	23.786	20.745	-3.04173
##	1711	3384	38.121	6.667	13.333	6.6667	0.000	9.333	9.33333
##	1712	11113	51.822	14.754	17.727	2.9732	13.115	12.727	-0.38748
##	1800	1469	72.703	20.000	32.432	12.4324	27.500	24.324	-3.17568
##	1801	26	73.077	NaN	NaN	NaN	NaN	NaN	NaN
##	1802	2383	64.373	25.714	20.202	-5.5123	25.714	26.263	0.54834
##	1803	5394	50.612	17.829	22.876	5.0464	19.380	27.451	8.07114
##	1804	4082	60.167	11.594	20.253	8.6590	15.942	18.987	3.04531
##	1900	4737	49.947	7.438	32.667	25.2287	10.744	26.000	15.25620
##	1901	6456	44.006	9.195	18.947	9.7520	12.644	20.000	7.35632
##	1902	8627	44.720	12.121	31.050	18.9290	9.091	16.895	7.80407
##	1903	1002	54.391	10.000	36.364	26.3636	30.000	12.121	-17.87879
##	1904	7444	60.962	18.405	40.244	21.8390	22.699	27.439	4.73964
##	1905	720	52.917	6.250	17.857	11.6071	0.000	17.857	17.85714

##	1906	6957	46.500	10.656	24.348	13.6921	9.016	17.391	8.37491
##	1907	7425	47.178	15.714	22.066	6.3514	12.857	16.432	3.57478
##	1908	6508	46.650	11.111	23.585	12.4738	11.111	17.925	6.81342
##	1909	3382	32.436	15.385	17.273	1.8881	7.692	19.091	11.39860
##	1910	6655	49.391	14.035	34.653	20.6184	10.526	25.743	15.21626
##	1911	4162	58.602	17.708	20.561	2.8524	21.875	14.019	-7.85631
##	1912	11132	32.716	17.797	18.248	0.4516	9.322	12.409	3.08673
##	1913	892	54.596	0.000	22.222	22.2222	8.333	13.889	5.55556
##	2000	3830	74.569	8.889	24.370	15.4809	17.778	28.571	10.79365
##	2001	2056	80.350	30.000	37.430	7.4302	35.000	31.285	-3.71508
##	2002	21243	74.420	11.615	29.492	17.8767	12.160	26.766	14.60609
##	2003	7280	70.549	10.417	19.703	9.2859	11.806	19.703	7.89705
##	2100	3326	55.292	15.385	25.658	10.2733	3.846	18.421	14.57490
##	2101	461	54.881	100.000	27.778	-72.2222	100.000	8.333	-91.66667
##	2102	6500	33.246	16.667	23.762	7.0957	11.111	12.871	1.76018
##	2103	4193	35.512	16.071	29.104	13.0330	12.500	17.164	4.66418
##	2104	2680	27.612	10.000	33.333	23.3333	10.000	20.430	10.43011
##	2105	6270	45.789	14.286	31.549	17.2636	16.327	21.127	4.80023
##	2200	11196	38.621	12.121	20.625	8.5038	10.909	15.625	4.71591
##	2201	1956	49.949	10.000	24.359	14.3590	20.000	16.667	-3.33333
##	2202	5061	34.183	6.977	17.021	10.0445	6.977	7.447	0.47006
##	2203	1713	43.900	0.000	33.333	33.3333	0.000	20.635	20.63492
##	2204	7824	47.444	14.493	31.803	17.3105	15.942	19.672	3.73010
##	2205	11245	40.507	3.614	24.167	20.5522	8.434	15.833	7.39960
##	2206	2092	22.945	0.000	8.333	8.3333	0.000	8.333	8.33333
##	2207	7197	36.696	8.974	24.242	15.2681	6.410	7.576	1.16550
##	2208	28245	31.046	8.293	12.432	4.1397	5.366	5.405	0.03955
##	2209	10084	36.265	12.500	20.074	7.5743	0.000	17.472	17.47212
##	2210	26843	22.177	4.688	16.486	11.7980	3.906	9.420	5.51404
##	2211	18782	22.048	7.547	17.419	9.8722	5.660	13.871	8.21059
##	2212	2864	36.767	7.895	14.815	6.9201	10.526	16.667	6.14035
##	2213	4562	46.405	22.500	31.609	9.1092	10.000	21.839	11.83908
##	2214	1195	59.331	14.286	45.455	31.1688	33.333	43.182	9.84848
##	2215	6916	42.206	7.407	24.793	17.3860	11.111	10.744	-0.36731
##	2216	2278	65.891	13.333	39.362	26.0284	13.333	37.234	23.90071
##	2300	10908	53.850	16.788	35.906	19.1177	16.058	24.161	8.10268
##	2301	4991	74.214	26.000	42.857	16.8571	25.000	49.143	24.14286
##	2302	1462	50.342	14.286	40.909	26.6234	14.286	22.727	8.44156
##	2303	11816	55.061	20.745	39.444	18.6998	20.745	27.778	7.03310
##	2304	8490	47.845	18.182	40.284	22.1025	13.131	24.171	11.03930
##	2305	5770	37.071	24.528	30.719	6.1907	18.868	24.183	5.31508
##	2306	2928	65.198	24.000	45.342	21.3416	40.000	27.329	-12.67081
##	2307	4050	48.691	23.810	44.776	20.9666	17.857	34.328	16.47122
##	2308	13295	67.198	21.000	39.903	18.9031	19.500	32.472	12.97173
##	2309	4631	64.263	24.561	36.078	11.5170	19.298	29.804	10.50568
##	2310	6987	40.332	21.538	36.453	14.9147	23.077	25.616	2.53884
##	2311	4399	41.032	22.500	36.296	13.7963	20.000	26.667	6.66667
##	2312	8584	42.556	14.151	30.055	15.9037	12.264	22.404	10.14022
##	2400	5907	58.202	23.729	44.545	20.8166	16.949	25.000	8.05085
##	2401	94	65.957	NaN	50.000	NaN	NaN	25.000	NaN

##	2402	4761	40.244	31.092	44.118	13.0252	12.605	25.000	12.39496
##	2403	17090	49.222	38.643	44.272	5.6294	15.044	26.316	11.27154
##	2404	10478	55.039	36.364	55.357	18.9935	14.050	33.036	18.98613
##	2405	4210	46.485	31.507	45.631	14.1242	21.918	24.272	2.35404
##	2406	4899	52.766	34.568	60.396	25.8281	13.580	33.663	20.08312
##	2500	24061	31.557	9.890	22.222	12.3321	8.242	14.762	6.52015
##	2501	1151	29.974	12.500	34.615	22.1154	12.500	38.462	25.96154
##	2502	4630	46.631	19.444	36.496	17.0519	11.111	21.898	10.78670
##	2503	10762	41.005	10.588	27.673	17.0847	7.843	13.208	5.36441
##	2504	19737	33.860	11.679	17.872	6.1935	5.839	12.766	6.92654
##	2505	21299	43.594	12.689	30.282	17.5923	9.091	16.901	7.81050
##	2506	10665	40.619	10.915	25.874	14.9586	8.451	12.587	4.13671
##	2507	6616	31.258	13.393	25.532	12.1391	11.607	18.085	6.47796
##	2508	14289	39.499	12.453	21.429	8.9757	8.679	13.839	5.16004
##	2600	8446	54.689	5.946	12.891	6.9447	6.486	12.500	6.01351
##	2601	1039	61.598	15.385	18.519	3.1339	0.000	11.111	11.11111
##	2602	2506	64.286	9.524	23.656	14.1321	7.143	18.280	11.13671
##	2603	2092	50.574	12.500	18.310	5.8099	12.500	19.718	7.21831
##	2604	15346	43.106	8.716	16.667	7.9511	7.798	11.559	3.76097
##	2605	3347	49.029	15.000	11.966	-3.0342	10.000	19.658	9.65812
##	2606	1549	48.612	11.111	19.403	8.2919	11.111	8.955	-2.15589
##	2607	1351	66.247	0.000	15.217	15.2174	0.000	4.348	4.34783
##	2608	1481	64.416	0.000	14.545	14.5455	9.091	9.091	0.00000
##	2609	886	71.445	0.000	17.391	17.3913	0.000	17.391	17.39130
##	2610	4793	41.060	6.410	6.667	0.2564	8.974	4.762	-4.21245
##	2611	8550	41.485	13.636	19.331	5.6945	13.636	15.613	1.97702
##	2612	1254	36.762	7.692	19.048	11.3553	7.692	14.286	6.59341
##	2613	10356	57.426	13.018	17.277	4.2597	12.426	16.230	3.80433
##	2614	4413	58.260	4.688	14.615	9.9279	6.250	13.077	6.82692
##	2700	56561	57.138	26.667	50.155	23.4886	21.844	38.121	16.27715
##	2701	11559	60.611	32.609	57.344	24.7354	29.891	39.437	9.54532
##	2702	2289	53.342	40.909	58.491	17.5815	31.818	32.075	0.25729
##	2703	8940	39.497	40.741	64.912	24.1715	20.370	48.246	27.87524
##	2704	1912	44.247	30.000	56.250	26.2500	13.333	34.375	21.04167
##	2705	21977	51.791	19.398	25.636	6.2380	8.696	14.873	6.17715
##	2706	7751	50.148	20.000	26.623	6.6234	8.000	24.026	16.02597
##	2707	1548	54.457	27.778	43.243	15.4655	22.222	43.243	21.02102
##	2708	7918	28.934	43.662	59.483	15.8208	36.620	56.897	20.27683
##	2709	291	87.629	25.000	77.778	52.7778	37.500	72.222	34.72222
##	2710	1230	47.480	47.826	51.852	4.0258	30.435	37.037	6.60225
##	2711	6144	38.525	17.073	20.792	3.7189	7.317	19.802	12.48491
##	2712	11581	40.964	27.692	47.598	19.9059	13.077	28.821	15.74404
##	2713	5910	54.941	37.288	51.724	14.4360	27.119	41.379	14.26067
##	2714	2759	72.852	34.694	65.934	31.2402	38.776	37.363	-1.41287
##	2715	7356	38.146	21.818	30.464	8.6454	5.455	18.543	13.08850
##	2716	7396	48.053	43.697	58.046	14.3485	25.210	43.103	17.89336
##	2717	4738	64.584	40.260	62.353	22.0932	32.468	50.000	17.53247
##	2718	2754	61.438	35.294	44.898	9.6038	35.294	36.054	0.76030
##	2719	12276	71.554	45.263	51.216	5.9526	35.263	46.353	11.08943
##	2720	8466	43.952	35.965	46.721	10.7564	23.684	29.508	5.82399

##	2721	2863	57.143	21.667	36.620	14.9531	6.667	16.901	10.23474
##	2722	2922	35.558	33.333	55.556	22.2222	25.926	13.333	-12.59259
##	2723	9483	48.360	36.364	47.034	10.6703	19.886	30.085	10.19838
##	2724	6400	41.875	20.732	42.735	22.0033	12.195	19.658	7.46300
##	2725	13598	43.264	34.615	52.011	17.3961	24.615	40.517	15.90186
##	2726	5959	39.168	35.714	47.222	11.5079	30.952	34.028	3.07540
##	2727	3781	61.359	19.048	34.940	15.8921	19.048	27.711	8.66322
##	2728	18915	50.040	31.915	44.052	12.1372	17.021	25.465	8.44341
##	2729	16012	38.234	39.916	72.901	32.9848	31.092	55.344	24.25107
##	2730	16934	40.859	44.324	46.632	2.3078	25.405	31.347	5.94174
##	2731	10772	48.013	24.224	39.841	15.6170	14.286	22.709	8.42345
##	2732	19165	47.435	25.926	25.706	-0.2201	22.222	20.654	-1.56843
##	2733	9013	29.624	19.048	43.182	24.1342	11.905	27.273	15.36797
##	2734	8585	54.083	24.204	49.296	25.0920	17.834	42.958	25.12335
##	2735	17079	51.039	43.891	60.586	16.6942	37.557	43.243	5.68668
##	2736	11806	49.500	24.031	50.992	26.9605	12.403	33.711	21.30795
##	2737	6704	58.696	27.941	38.136	10.1944	8.824	22.034	13.21037
##	2738	29455	63.239	34.483	61.809	27.3263	33.266	46.935	13.66895
##	2739	27990	57.667	41.814	58.866	17.0524	34.509	49.072	14.56335
##	2740	10269	45.486	25.234	43.774	18.5399	13.084	27.547	14.46306
##	2741	17630	38.877	28.125	37.632	9.5066	20.000	25.000	5.00000
##	2742	5716	56.456	62.651	57.426	-5.2249	60.241	49.010	-11.23106
##	2743	5721	42.423	54.737	62.222	7.4854	40.000	47.778	7.77778
##	2744	434	59.217	33.333	NaN	NaN	21.930	NaN	NaN
##	2745	4584	43.216	34.884	50.000	15.1163	20.930	34.848	13.91825
##	2746	36275	34.768	17.188	23.631	6.4432	9.375	15.493	6.11796
##	2747	4195	46.675	27.451	49.153	21.7016	23.529	20.339	-3.19043
##	2748	5990	38.664	25.806	32.090	6.2831	29.032	18.657	-10.37554
##	2800	12948	56.333	30.870	42.319	11.4493	16.522	24.928	8.40580
##	2801	680	65.882	12.500	40.000	27.5000	12.500	26.667	14.16667
##	2802	4579	63.791	37.647	45.638	7.9905	30.588	40.940	10.35136
##	2803	2341	65.442	31.579	57.983	26.4042	21.053	39.496	18.44317
##	2804	7336	51.309	31.092	44.279	13.1862	13.445	33.333	19.88796
##	2805	7364	62.778	38.272	45.779	7.5076	30.864	31.169	0.30463
##	2806	1479	66.058	42.308	57.143	14.8352	23.077	60.000	36.92308
##	2807	903	39.313	28.571	45.455	16.8831	0.000	31.818	31.81818
##	2808	9609	52.003	37.113	53.875	16.7611	21.649	32.841	11.19184
##	2809	6733	51.300	20.423	42.143	21.7203	11.972	22.143	10.17103
##	2900	11256	67.697	54.380	69.286	14.9062	55.839	67.857	12.01773
##	2901	854	34.543	NaN	50.000	NaN	NaN	50.000	NaN
##	2902	3861	70.060	26.471	63.636	37.1658	11.765	51.240	39.47496
##	2903	95	81.053	0.000	75.000	75.0000	14.286	75.000	60.71429
##	2904	404	83.416	NaN	63.636	NaN	NaN	45.455	NaN
##	2905	2346	82.907	NaN	72.611	NaN	NaN	68.153	NaN
##	2906	835	68.024	79.167	40.000	-39.1667	75.000	26.667	-48.33333
##	2907	1034	70.019	14.286	42.105	27.8195	14.286	31.579	17.29323
##	2908	649	29.430	NaN	70.000	NaN	NaN	65.000	NaN
##	2909	1641	69.287	60.000	77.941	17.9412	20.000	63.235	43.23529
##	2910	2204	71.370	36.957	44.211	7.2540	34.783	40.000	5.21739
##	2911	2050	65.756	60.000	29.268	-30.7317	33.333	36.585	3.25203

##	2912	442	78.054	84.615	71.429	-13.1868	46.154	92.857	46.70330
##	2913	1926	77.259	80.000	90.722	10.7216	60.000	80.412	20.41237
##	2914	1412	78.045	50.000	65.625	15.6250	50.000	75.000	25.00000
##	2915	15	86.667	NaN	100.000	NaN	NaN	100.000	NaN
##	2916	6620	51.662	40.127	65.877	25.7494	41.401	42.654	1.25275
##	2917	852	76.643	100.000	60.870	-39.1304	100.000	67.391	-32.60870
##	2918	0	NA	NA	NA	NA	NA	NA	NA
##	2919	1151	82.798	72.340	71.429	-0.9119	70.213	69.048	-1.16515
##	2920	72	69.444	0.000	NaN	NaN	50.000	NaN	NaN
##	2921	3453	61.454	54.545	64.865	10.3194	59.091	50.000	-9.09091
##	2922	680	87.794	62.069	55.000	-7.0690	62.069	60.000	-2.06897
##	2923	488	80.943	NaN	50.000	NaN	NaN	41.667	NaN
##	3000	1126	59.947	42.857	42.222	-0.6349	0.000	35.556	35.55556
##	3001	329	76.900	0.000	20.000	20.0000	12.500	40.000	27.50000
##	3002	8957	68.628	16.418	34.715	18.2971	11.940	14.508	2.56747
##	3003	8398	56.811	33.908	40.000	6.0920	17.241	27.442	10.20048
##	3004	16926	51.755	37.581	45.290	7.7089	11.663	24.638	12.97461
##	3005	5047	50.248	31.967	41.284	9.3172	16.393	33.028	16.63408
##	3100	21427	39.758	15.000	21.595	6.5953	20.000	16.537	-3.46304
##	3101	7944	33.686	9.375	10.625	1.2500	9.375	9.375	0.00000
##	3102	4529	36.321	7.576	21.053	13.4769	7.576	14.474	6.89793
##	3103	8626	32.205	19.192	19.577	0.3848	10.101	15.873	5.77201
##	3104	37758	29.117	12.577	19.601	7.0246	6.442	10.797	4.35562
##	3105	7800	32.154	12.857	18.182	5.3247	8.571	12.299	3.72804
##	3106	9134	42.884	11.236	18.182	6.9459	8.989	9.091	0.10215
##	3107	12929	35.857	9.259	15.625	6.3657	7.407	11.648	4.24032
##	3108	2118	37.441	19.048	35.294	16.2465	23.810	20.588	-3.22129
##	3109	5829	40.847	6.173	8.209	2.0361	9.877	7.463	-2.41386
##	3110	6143	26.258	12.903	30.645	17.7419	8.065	12.903	4.83871
##	3200	9918	75.227	34.752	52.083	17.3316	33.333	44.792	11.45833
##	3201	1286	76.205	30.303	57.143	26.8398	27.273	50.000	22.72727
##	3202	8148	71.244	34.659	55.474	20.8154	37.500	45.742	8.24209
##	3203	11163	77.076	33.953	66.071	32.1179	40.465	51.389	10.92377
##	3204	13412	75.514	46.689	66.607	19.9187	44.040	55.595	11.55529
##	3205	11686	71.128	25.333	51.483	26.1497	30.667	39.831	9.16384
##	3206	3941	66.176	30.612	54.348	23.7356	22.449	39.130	16.68146
##	3207	8339	79.770	44.079	47.712	3.6335	40.132	44.662	4.53073
##	3300	8548	81.902	22.759	44.474	21.7151	24.828	43.421	18.59347
##	3301	8785	80.968	31.771	49.774	18.0029	32.812	47.285	14.47257
##	3302	5884	74.354	26.316	35.622	9.3065	29.825	32.189	2.36428
##	3303	11788	77.833	27.797	41.364	13.5670	28.814	37.273	8.45917
##	3304	39517	79.619	32.558	56.367	23.8088	31.190	50.754	19.56422
##	3305	32930	75.840	22.581	37.520	14.9397	23.349	35.400	12.05098
##	3306	12628	78.659	41.538	61.316	19.7778	43.077	57.303	14.22645
##	3307	2193	67.487	22.388	51.282	28.8940	19.403	24.359	4.95599
##	3308	20938	81.569	33.043	42.295	9.2510	33.043	43.836	10.79214
##	3309	6815	79.574	32.941	37.696	4.7552	30.980	45.026	14.04579
##	3310	14862	80.057	41.275	54.958	13.6828	38.926	52.269	13.34273
##	3311	2020	75.891	30.000	35.294	5.2941	10.000	38.235	28.23529
##	3312	41076	84.061	29.249	42.448	13.1989	27.536	40.712	13.17557

```

## 3313 3573      65.939 18.310 32.215 13.9049 15.493 24.832 9.33926
## 3314 7002      80.363 32.979 45.365 12.3862 31.915 43.787 11.87209
## 3315 7044      84.696 30.000 48.207 18.2072 30.000 50.398 20.39841
## 3316 17203     85.165 31.081 43.187 12.1062 31.081 43.665 12.58426
## 3317 2426      76.298 37.500 62.037 24.5370 34.375 57.407 23.03241
## 3318 4046      82.921 82.022 73.029 -8.9934 77.528 71.369 -6.15880
## 3319 2103      81.978 48.000 60.484 12.4839 42.000 54.032 12.03226
## 3320 17209     82.503 21.717 33.404 11.6869 21.212 33.934 12.72213
## 3321 5867      82.785 25.444 42.982 17.5387 21.893 39.912 18.01879
## 3322 4979      77.606 26.852 31.884 5.0322 25.000 36.715 11.71498
## 3400 9349      39.972 32.743 60.902 28.1589 24.779 45.113 20.33402
## 3401 306       28.758 80.000 70.000 -10.0000 40.000 40.000 0.00000
## 3402 1170      7.863 80.000 70.000 -10.0000 80.000 90.000 10.00000
## 3403 1308      36.774 9.091 70.588 61.4973 27.273 50.000 22.72727
## 3404 1289      17.921 0.000 35.714 35.7143 0.000 35.714 35.71429
## 3500 6737      24.566 27.273 46.739 19.4664 27.273 27.174 -0.09881
## 3501 161       22.981      NaN 25.000      NaN      NaN 25.000      NaN
## 3502 0          NA      NA      NA      NA      NA      NA      NA
## 3503 9          22.222      NaN      NaN      NaN      NaN      NaN      NaN
## 3504 2693      31.341 5.882 40.625 34.7426 5.882 12.500 6.61765
## 3505 815       44.540 66.667 80.000 13.3333 66.667 53.333 -13.33333
## 3506 444       29.955 11.111 20.000 8.8889 0.000 40.000 40.00000
## 3600 1491      66.130 44.828 56.000 11.1724 48.276 64.000 15.72414
## 3601 1090      74.495 66.667 57.143 -9.5238 66.667 47.619 -19.04762
## 3602 358       73.464      NaN 53.333      NaN      NaN 53.333      NaN
## 3603 388       44.330 21.429 25.000 3.5714 21.429 37.500 16.07143
## 3604 49        59.184      NaN 83.333      NaN      NaN 33.333      NaN
## 3605 675       61.778 0.000 30.952 30.9524 0.000 35.714 35.71429
## 3606 0          NA      NA      NA      NA      NA      NA      NA
## 3607 558       68.100 11.111 25.000 13.8889 33.333 18.750 -14.58333
## 3608 19       100.000      NaN 75.000      NaN      NaN 75.000      NaN
## 3609 805       71.429 81.818 84.615 2.7972 75.758 61.538 -14.21911
## 3610 793       62.169 11.765 45.000 33.2353 5.882 25.000 19.11765
## 3611 962       68.711 82.353 53.488 -28.8646 35.294 58.140 22.84542
## 3612 8886      67.634 55.063 37.650 -17.4134 54.430 31.175 -23.25532
## 3613 459       35.294 0.000 40.000 40.0000 0.000 0.000 0.00000
## 3614 4339      37.705 35.714 36.620 0.9054 35.714 29.577 -6.13682
## 3615 0          NA      NA      NA      NA      NA      NA      NA
## 3616 2523      70.947 69.643 66.265 -3.3778 58.929 54.217 -4.71170

```

```

print(paste("Stats just for gendered aticles from 1 country 1996 -",
LastYearForCitationAnalysis))

```

```

## [1] "Stats just for gendered aticles from 1 country 1996 - 2012"

```

```

print(BasicStats2)

```

```

##      1CtyMFto14 Fem1All Fem1 96 Fem1 14      Chg1 FemNA11 FemN96 FemN14
## 1000      3074 23.780 14.857 28.444 13.58730 15.355 9.714 19.111
## 1100      5488 33.127 23.348 43.189 19.84100 20.627 18.943 26.927
## 1101      903 26.578 25.581 30.000 4.41860 16.058 13.953 24.000

```

## 1102	1743	30.752	30.000	29.915	-0.08547	16.007	12.500	17.949
## 1103	4109	33.414	24.878	44.966	20.08839	22.536	21.463	29.866
## 1104	3251	30.145	22.905	43.961	21.05632	20.517	21.229	22.705
## 1105	9271	30.115	22.042	38.235	16.19353	19.243	19.258	22.409
## 1106	2835	39.577	38.693	37.158	-1.53500	27.972	28.643	27.869
## 1107	882	22.449	11.364	26.087	14.72332	15.420	9.091	20.290
## 1108	392	32.908	22.449	40.000	17.55102	16.327	6.122	25.000
## 1109	1614	33.643	30.097	37.391	7.29422	17.224	11.650	19.130
## 1110	3949	36.693	30.363	44.344	13.98086	20.410	18.482	22.624
## 1111	960	29.375	21.569	50.000	28.43137	14.375	11.765	20.968
## 1200	1155	33.247	26.667	42.105	15.43860	33.160	33.333	42.105
## 1201	8287	40.485	33.430	46.631	13.20121	37.118	33.721	41.612
## 1202	11730	31.790	21.448	36.046	14.59784	31.279	22.520	35.883
## 1203	7277	44.167	39.373	48.518	9.14512	42.971	37.282	47.894
## 1204	2356	33.786	28.750	38.397	9.64662	29.966	35.000	36.709
## 1205	779	34.018	17.073	51.163	34.08962	34.146	17.073	50.000
## 1206	275	40.000	0.000	44.828	44.82759	38.182	0.000	48.276
## 1207	3369	35.945	25.758	39.057	13.29966	33.422	23.485	36.364
## 1208	5852	41.029	33.000	44.514	11.51411	41.217	33.000	44.044
## 1209	194	46.392	0.000	70.000	70.00000	45.361	0.000	70.000
## 1210	940	32.660	23.684	39.130	15.44622	29.894	23.684	33.333
## 1211	6279	22.217	15.487	27.052	11.56576	21.659	14.159	27.187
## 1212	3161	25.435	27.885	28.483	0.59836	25.150	24.038	27.554
## 1213	3679	43.354	35.870	48.952	13.08282	42.974	35.870	47.619
## 1300	6786	32.862	20.621	41.712	21.09074	19.879	13.842	26.685
## 1301	16	18.750	NaN	0.000	NaN	18.750	NaN	33.333
## 1302	857	41.424	35.000	56.842	21.84211	27.655	25.000	27.368
## 1303	20604	31.149	26.808	36.075	9.26688	14.978	12.637	16.355
## 1304	4195	32.253	23.984	31.579	7.59521	15.232	13.008	17.982
## 1305	3927	34.097	30.000	35.959	5.95890	18.309	12.000	21.575
## 1306	3990	44.762	41.111	49.259	8.14815	25.764	21.111	30.741
## 1307	10590	43.088	34.265	47.874	13.60960	20.633	16.765	23.475
## 1308	3663	36.746	24.859	40.278	15.41902	24.952	18.644	24.074
## 1309	2536	47.674	50.360	51.786	1.42600	29.692	28.777	32.143
## 1310	2386	42.037	29.927	51.656	21.72862	22.171	15.328	26.490
## 1311	8973	39.452	35.447	41.278	5.83091	20.840	20.650	22.107
## 1312	16421	37.945	31.480	42.621	11.14137	18.592	14.300	20.661
## 1313	3346	33.144	23.669	46.512	22.84299	19.605	14.793	24.031
## 1314	5198	35.379	31.118	41.994	10.87613	19.104	12.387	26.284
## 1315	2554	31.676	27.815	38.333	10.51876	17.267	15.232	22.500
## 1400	4132	25.920	15.385	28.740	13.35554	25.895	18.376	27.165
## 1401	2016	31.349	30.357	33.607	3.24941	30.357	31.250	31.148
## 1402	2244	26.872	18.919	35.484	16.56495	23.797	16.216	30.323
## 1403	4433	25.378	22.000	29.744	7.74359	25.761	23.333	27.692
## 1404	981	22.426	13.699	36.364	22.66501	22.528	15.068	31.818
## 1405	4332	25.092	16.183	26.592	10.40919	24.307	17.842	27.341
## 1406	2682	32.662	27.642	33.051	5.40857	31.805	27.642	31.356
## 1407	4271	33.646	25.592	40.483	14.89097	33.318	26.540	44.109
## 1408	7628	26.796	22.345	25.996	3.65107	27.019	21.903	27.514
## 1409	2206	34.270	35.165	38.554	3.38938	34.451	30.769	42.169

## 1410	867	26.298	18.519	25.532	7.01340	25.029	25.926	23.404
## 1500	2658	17.306	15.672	23.320	7.64852	11.550	11.194	12.253
## 1501	131	27.481	0.000	33.333	33.33333	27.481	0.000	50.000
## 1502	1934	28.697	22.078	25.131	3.05297	16.649	14.286	13.089
## 1503	4922	21.312	11.301	27.273	15.97136	11.113	7.534	9.971
## 1504	83	31.325	50.000	33.333	-16.66667	14.458	50.000	33.333
## 1505	1121	25.156	17.073	26.667	9.59350	12.043	12.195	6.667
## 1506	155	23.871	0.000	40.000	40.00000	14.839	0.000	0.000
## 1507	306	14.379	13.333	20.000	6.66667	10.131	6.667	0.000
## 1508	616	18.344	10.000	12.500	2.50000	13.636	3.333	18.750
## 1600	13405	21.037	14.319	26.797	12.47761	12.160	12.922	13.276
## 1601	303	28.053	14.286	31.579	17.29323	21.782	7.143	42.105
## 1602	3501	29.363	19.028	30.901	11.87295	14.596	12.955	17.597
## 1603	1567	26.101	10.000	32.773	22.77311	8.679	4.286	6.723
## 1604	4247	21.733	14.844	32.524	17.68052	13.139	8.203	14.078
## 1605	10203	20.269	15.721	26.975	11.25372	12.065	10.049	11.561
## 1606	9488	19.709	14.136	21.832	7.69623	10.371	7.853	10.916
## 1607	2607	27.043	20.134	30.488	10.35358	12.965	6.040	12.805
## 1700	1678	17.044	14.286	19.008	4.72255	15.852	12.088	16.529
## 1701	195	16.410	NaN	23.529	NaN	17.436	NaN	17.647
## 1702	2802	20.414	17.213	21.101	3.88780	16.774	16.393	16.514
## 1703	1686	14.947	9.091	20.130	11.03896	12.218	6.818	14.286
## 1704	722	15.235	9.091	17.857	8.76623	13.712	9.091	25.000
## 1705	2098	16.873	8.861	23.741	14.88025	15.014	8.861	21.583
## 1706	5578	20.186	14.904	23.913	9.00920	17.300	19.231	19.746
## 1707	909	13.201	4.545	20.000	15.45455	10.781	2.273	22.857
## 1708	1158	14.421	14.754	21.311	6.55738	12.176	11.475	11.475
## 1709	1299	28.329	26.531	24.528	-2.00231	24.018	28.571	22.642
## 1710	3202	27.046	26.699	22.086	-4.61314	23.923	23.786	19.632
## 1711	842	10.808	6.667	15.385	8.71795	8.314	0.000	13.846
## 1712	4103	15.135	14.754	13.410	-1.34414	13.137	13.115	11.494
## 1800	693	26.407	20.000	33.333	13.33333	24.242	27.500	20.370
## 1801	11	9.091	NaN	0.000	NaN	18.182	NaN	0.000
## 1802	1012	22.036	25.714	23.944	-1.77062	19.170	25.714	26.761
## 1803	1875	18.027	17.829	23.364	5.53503	17.227	19.380	15.888
## 1804	1528	14.725	11.594	18.898	7.30343	12.107	15.942	13.386
## 1900	1579	20.393	7.438	30.000	22.56198	18.303	10.744	28.182
## 1901	2063	19.486	9.195	30.469	21.27335	14.978	12.644	13.281
## 1902	2548	19.702	12.121	34.595	22.47338	14.089	9.091	22.162
## 1903	316	25.000	10.000	48.889	38.88889	20.253	30.000	33.333
## 1904	3187	23.753	18.405	31.839	13.43366	21.556	22.699	26.906
## 1905	240	11.667	6.250	15.789	9.53947	7.917	0.000	10.526
## 1906	2262	20.866	10.656	30.000	19.34426	12.069	9.016	12.667
## 1907	2365	16.871	15.714	21.656	5.94177	12.474	12.857	13.376
## 1908	2115	18.913	11.111	23.529	12.41830	12.104	11.111	12.500
## 1909	583	13.722	15.385	15.217	-0.16722	10.463	7.692	8.696
## 1910	2435	24.066	14.035	33.784	19.74870	16.509	10.526	18.919
## 1911	1746	20.046	17.708	26.531	8.82228	16.208	21.875	16.327
## 1912	2492	17.014	17.797	17.341	-0.45557	10.273	9.322	12.139
## 1913	329	19.149	0.000	26.087	26.08696	13.982	8.333	21.739



## 2000	1800	20.222	8.889	28.221	19.33197	21.167	17.778	25.767
## 2001	851	24.559	30.000	28.155	-1.84466	22.562	35.000	25.243
## 2002	10875	20.460	11.615	23.852	12.23694	20.028	12.160	23.292
## 2003	3474	19.891	10.417	22.932	12.51566	18.653	11.806	18.797
## 2100	935	21.283	15.385	24.306	8.92094	14.652	3.846	17.361
## 2101	103	14.563	100.000	25.000	-75.00000	13.592	100.000	18.750
## 2102	1041	13.064	16.667	16.964	0.29762	7.397	11.111	6.250
## 2103	800	16.125	16.071	22.973	6.90154	8.625	12.500	8.108
## 2104	327	11.927	10.000	21.622	11.62162	7.034	10.000	5.405
## 2105	1091	19.798	14.286	17.610	3.32435	15.582	16.327	17.610
## 2200	3104	17.622	12.121	17.188	5.06629	13.724	10.909	14.062
## 2201	588	20.068	10.000	32.759	22.75862	18.197	20.000	18.966
## 2202	1066	8.161	6.977	11.881	4.90444	6.379	6.977	6.931
## 2203	411	9.976	0.000	7.843	7.84314	9.246	0.000	7.843
## 2204	1973	28.789	14.493	34.454	19.96103	18.196	15.942	17.647
## 2205	2604	11.290	3.614	17.037	13.42258	11.790	8.434	14.074
## 2206	329	11.550	0.000	18.182	18.18182	6.687	0.000	9.091
## 2207	1697	10.194	8.974	6.429	-2.54579	9.664	6.410	5.714
## 2208	5831	10.153	8.293	10.824	2.53085	7.237	5.366	7.294
## 2209	2233	16.256	12.500	15.676	3.17568	15.584	0.000	14.054
## 2210	3104	10.664	4.688	12.707	8.01968	8.409	3.906	6.630
## 2211	2421	16.274	7.547	15.254	7.70707	9.624	5.660	11.864
## 2212	667	11.844	7.895	18.519	10.62378	9.595	10.526	14.815
## 2213	1187	20.640	22.500	19.853	-2.64706	16.933	10.000	13.971
## 2214	454	23.128	14.286	42.857	28.57143	23.128	33.333	35.714
## 2215	1709	13.868	7.407	23.418	16.01031	13.985	11.111	17.089
## 2216	890	20.112	13.333	32.857	19.52381	18.764	13.333	30.000
## 2300	4060	24.951	16.788	30.272	13.48379	19.729	16.058	26.531
## 2301	2524	27.892	26.000	30.126	4.12552	28.566	25.000	31.799
## 2302	449	26.503	14.286	28.814	14.52785	14.254	14.286	6.780
## 2303	4278	28.752	20.745	34.862	14.11770	18.700	20.745	20.183
## 2304	2713	29.451	18.182	31.020	12.83859	16.955	13.131	15.918
## 2305	1300	22.231	24.528	26.515	1.98685	16.385	18.868	19.697
## 2306	1016	23.130	24.000	31.132	7.13208	19.882	40.000	24.528
## 2307	1264	39.399	23.810	43.836	20.02609	20.965	17.857	23.288
## 2308	5513	28.605	21.000	31.369	10.36882	26.410	19.500	26.806
## 2309	1799	28.460	24.561	34.483	9.92136	18.621	19.298	22.069
## 2310	1707	29.643	21.538	32.168	10.62937	18.805	23.077	20.979
## 2311	1065	28.451	22.500	24.742	2.24227	18.216	20.000	20.619
## 2312	2400	23.917	14.151	33.529	19.37847	16.417	12.264	21.765
## 2400	1985	31.788	23.729	38.830	15.10097	18.741	16.949	23.404
## 2401	21	47.619	NaN	66.667	NaN	33.333	NaN	0.000
## 2402	1525	37.180	31.092	35.821	4.72846	16.197	12.605	19.403
## 2403	6297	44.450	38.643	43.523	4.88025	22.979	15.044	26.166
## 2404	4354	41.984	36.364	40.164	3.80030	16.008	14.050	22.131
## 2405	1348	40.727	31.507	37.838	6.33099	23.813	21.918	22.523
## 2406	1901	42.294	34.568	40.299	5.73061	19.726	13.580	21.642
## 2500	4361	19.445	9.890	23.919	14.02847	10.731	8.242	12.468
## 2501	247	21.862	12.500	25.000	12.50000	20.243	12.500	37.500
## 2502	1249	33.707	19.444	32.026	12.58170	17.454	11.111	18.954

## 2503	3327	20.800	10.588	29.064	18.47580	9.889	7.843	11.823
## 2504	4554	18.621	11.679	22.254	10.57469	9.486	5.839	12.113
## 2505	7360	20.367	12.689	27.326	14.63619	11.359	9.091	14.826
## 2506	3257	18.023	10.915	30.270	19.35478	9.917	8.451	13.514
## 2507	1428	22.899	13.393	34.146	20.75348	12.535	11.607	17.073
## 2508	4085	20.294	12.453	25.781	13.32842	10.086	8.679	13.672
## 2600	2985	12.797	5.946	14.118	8.17170	11.323	6.486	11.373
## 2601	400	15.250	15.385	18.919	3.53430	13.750	0.000	18.919
## 2602	1069	12.535	9.524	14.773	5.24892	12.535	7.143	17.045
## 2603	636	10.220	12.500	4.688	-7.81250	9.119	12.500	3.125
## 2604	4280	12.360	8.716	14.835	6.11957	9.977	7.798	11.264
## 2605	1018	12.672	15.000	15.385	0.38462	8.939	10.000	14.423
## 2606	436	9.862	11.111	5.405	-5.70571	11.239	11.111	16.216
## 2607	572	10.839	0.000	11.905	11.90476	9.965	0.000	4.762
## 2608	541	12.015	0.000	10.169	10.16949	10.351	9.091	8.475
## 2609	427	9.368	0.000	8.571	8.57143	6.557	0.000	2.857
## 2610	1415	10.883	6.410	10.127	3.71633	8.410	8.974	12.658
## 2611	2035	15.381	13.636	17.300	3.66321	11.400	13.636	14.768
## 2612	269	9.294	7.692	10.714	3.02198	8.178	7.692	10.714
## 2613	3675	14.231	13.018	15.493	2.47521	11.102	12.426	13.239
## 2614	1784	12.892	4.688	10.256	5.56891	11.547	6.250	7.692
## 2700	19714	37.902	26.667	42.191	15.52459	30.461	21.844	32.436
## 2701	4117	45.470	32.609	46.770	14.16133	35.511	29.891	38.243
## 2702	843	44.840	40.909	33.333	-7.57576	30.724	31.818	14.493
## 2703	2453	37.016	40.741	37.021	-3.71946	29.189	20.370	25.532
## 2704	621	37.198	30.000	36.957	6.95652	17.713	13.333	17.391
## 2705	7970	22.321	19.398	28.007	8.60888	10.903	8.696	10.997
## 2706	2616	29.128	20.000	28.780	8.78049	20.910	8.000	23.415
## 2707	596	46.980	27.778	53.061	25.28345	39.933	22.222	30.612
## 2708	1531	48.726	43.662	57.000	13.33803	32.724	36.620	49.000
## 2709	178	50.000	25.000	42.857	17.85714	48.876	37.500	57.143
## 2710	432	54.167	47.826	70.588	22.76215	31.481	30.435	32.353
## 2711	1556	24.293	17.073	28.906	11.83308	18.316	7.317	19.531
## 2712	3124	39.885	27.692	52.490	24.79811	21.447	13.077	28.352
## 2713	2017	42.390	37.288	45.833	8.54520	30.937	27.119	30.729
## 2714	1320	46.667	34.694	54.615	19.92151	33.409	38.776	35.385
## 2715	1928	24.585	21.818	31.646	9.82739	10.010	5.455	17.089
## 2716	2410	46.473	43.697	55.851	12.15358	28.299	25.210	31.383
## 2717	1934	46.536	40.260	63.682	23.42185	35.160	32.468	45.274
## 2718	905	41.436	35.294	41.584	6.29004	35.249	35.294	32.673
## 2719	4898	46.019	45.263	47.036	1.77242	38.710	35.263	42.292
## 2720	2772	41.089	35.965	36.517	0.55194	22.835	23.684	23.034
## 2721	1143	24.934	21.667	38.889	17.22222	8.924	6.667	22.222
## 2722	684	37.573	33.333	36.066	2.73224	24.123	25.926	24.590
## 2723	3196	43.523	36.364	47.321	10.95779	24.499	19.886	30.357
## 2724	1853	34.808	20.732	47.015	26.28322	18.511	12.195	27.612
## 2725	3708	43.743	34.615	47.484	12.86889	26.241	24.615	34.277
## 2726	1559	42.335	35.714	48.148	12.43386	23.541	30.952	30.556
## 2727	1617	32.035	19.048	39.850	20.80201	16.450	19.048	15.038
## 2728	6101	39.780	31.915	45.766	13.85087	23.652	17.021	23.964

## 2729	4470	51.409	39.916	62.357	22.44145	35.660	31.092	46.008
## 2730	4715	44.284	44.324	43.360	-0.96389	26.426	25.405	28.726
## 2731	3640	32.060	24.224	30.916	6.69243	20.549	14.286	24.427
## 2732	5197	26.477	25.926	24.132	-1.79430	19.800	22.222	17.916
## 2733	1874	28.068	19.048	26.866	7.81805	18.570	11.905	22.388
## 2734	2827	41.210	24.204	46.795	22.59105	32.826	17.834	41.026
## 2735	5863	49.019	43.891	53.463	9.57180	35.869	37.557	36.364
## 2736	3296	39.047	24.031	42.857	18.82614	22.209	12.403	27.083
## 2737	2554	33.125	27.941	39.556	11.61438	17.306	8.824	24.444
## 2738	11932	46.623	34.483	53.838	19.35573	37.647	33.266	41.974
## 2739	10387	49.735	41.814	53.245	11.43159	39.135	34.509	41.947
## 2740	2990	28.629	25.234	35.017	9.78319	15.017	13.084	18.182
## 2741	4532	31.642	28.125	36.677	8.55212	19.550	20.000	24.138
## 2742	2137	58.493	62.651	51.813	-10.83713	50.866	60.241	45.078
## 2743	1782	54.153	54.737	64.423	9.68623	38.215	40.000	42.308
## 2744	257	29.572	33.333	NaN	NaN	22.957	21.930	NaN
## 2745	1211	50.867	34.884	52.000	17.11628	27.746	20.930	32.000
## 2746	8301	18.070	17.188	21.352	4.16481	10.216	9.375	12.100
## 2747	1548	28.747	27.451	32.184	4.73293	14.341	23.529	16.092
## 2748	1487	24.882	25.806	24.194	-1.61290	13.853	29.032	12.097
## 2800	5088	37.048	30.870	39.733	8.86377	20.342	16.522	23.467
## 2801	221	34.842	12.500	35.897	23.39744	24.887	12.500	33.333
## 2802	1886	42.683	37.647	46.528	8.88072	30.064	30.588	36.111
## 2803	820	43.902	31.579	49.462	17.88342	32.195	21.053	34.409
## 2804	2485	41.610	31.092	41.256	10.16317	21.690	13.445	26.457
## 2805	2861	39.427	38.272	47.636	9.36476	26.145	30.864	32.000
## 2806	678	52.065	42.308	47.059	4.75113	34.218	23.077	27.451
## 2807	215	51.628	28.571	45.455	16.88312	33.023	0.000	36.364
## 2808	3159	40.551	37.113	50.719	13.60602	24.343	21.649	28.058
## 2809	2482	30.379	20.423	29.167	8.74413	19.984	11.972	26.042
## 2900	5878	60.514	54.380	66.129	11.74947	56.618	55.839	60.753
## 2901	140	65.000	NaN	40.000	NaN	57.143	NaN	60.000
## 2902	1756	61.219	26.471	68.519	42.04793	50.399	11.765	63.580
## 2903	56	35.714	0.000	100.000	100.00000	14.286	14.286	0.000
## 2904	221	64.706	NaN	53.333	NaN	55.204	NaN	66.667
## 2905	970	68.041	NaN	69.565	NaN	68.969	NaN	73.188
## 2906	342	65.789	79.167	56.667	-22.50000	57.602	75.000	43.333
## 2907	494	47.166	14.286	65.517	51.23153	43.927	14.286	41.379
## 2908	116	75.862	NaN	64.286	NaN	73.276	NaN	64.286
## 2909	642	58.723	60.000	68.889	8.88889	52.025	20.000	61.111
## 2910	939	40.788	36.957	35.714	-1.24224	38.232	34.783	44.898
## 2911	1049	47.092	60.000	33.871	-26.12903	41.087	33.333	35.484
## 2912	251	63.347	84.615	43.478	-41.13712	53.386	46.154	43.478
## 2913	886	87.810	80.000	91.262	11.26214	81.490	60.000	93.204
## 2914	690	67.246	50.000	60.417	10.41667	62.754	50.000	58.333
## 2915	9	66.667	NaN	NaN	NaN	66.667	NaN	NaN
## 2916	2185	53.547	40.127	52.514	12.38658	39.039	41.401	40.782
## 2917	431	75.406	100.000	70.455	-29.54545	72.158	100.000	63.636
## 2918	0	NA	NA	NA	NA	NA	NA	NA
## 2919	655	68.092	72.340	77.419	5.07893	65.496	70.213	80.645

## 2920	42	35.714	0.000	0.000	0.00000	26.190	50.000	0.000
## 2921	1259	48.451	54.545	58.389	3.84381	47.101	59.091	54.362
## 2922	410	64.390	62.069	50.000	-12.06897	62.195	62.069	55.000
## 2923	171	52.047	NaN	32.258	NaN	49.123	NaN	41.935
## 3000	313	38.019	42.857	40.000	-2.85714	21.406	0.000	23.333
## 3001	175	20.571	0.000	20.000	20.00000	19.429	12.500	30.000
## 3002	4971	20.439	16.418	28.287	11.86894	14.021	11.940	18.327
## 3003	3467	35.131	33.908	38.565	4.65698	23.363	17.241	21.076
## 3004	6694	38.169	37.581	42.410	4.82865	20.421	11.663	23.133
## 3005	1765	40.113	31.967	48.113	16.14599	24.533	16.393	32.075
## 3100	5410	13.604	15.000	16.880	1.87980	8.706	20.000	12.020
## 3101	1822	12.239	9.375	11.321	1.94575	9.440	9.375	12.264
## 3102	1143	18.023	7.576	22.785	15.20905	11.286	7.576	16.456
## 3103	1741	15.910	19.192	11.565	-7.62729	11.086	10.101	12.925
## 3104	7253	18.282	12.577	19.343	6.76638	9.389	6.442	9.307
## 3105	1596	12.719	12.857	8.081	-4.77633	8.647	8.571	11.111
## 3106	2878	10.702	11.236	11.446	0.20983	8.965	8.989	13.855
## 3107	2800	14.679	9.259	15.038	5.77833	7.714	7.407	12.406
## 3108	551	23.412	19.048	27.273	8.22511	13.975	23.810	15.909
## 3109	1604	9.850	6.173	8.943	2.77025	7.481	9.877	9.756
## 3110	1131	20.336	12.903	29.885	16.98183	9.107	8.065	10.345
## 3200	4976	39.871	34.752	45.550	10.79797	36.857	33.333	41.885
## 3201	650	43.231	30.303	43.548	13.24536	38.154	27.273	38.710
## 3202	3447	46.040	34.659	52.647	17.98797	39.252	37.500	44.412
## 3203	5208	50.384	33.953	60.861	26.90757	42.761	40.465	48.924
## 3204	6630	51.916	46.689	59.007	12.31861	45.897	44.040	51.471
## 3205	5499	39.462	25.333	48.307	22.97366	33.661	30.667	44.695
## 3206	1689	40.320	30.612	56.911	26.29832	33.748	22.449	44.715
## 3207	4013	47.396	44.079	56.934	12.85536	45.203	40.132	52.311
## 3300	4399	35.417	22.759	39.547	16.78798	34.212	24.828	40.302
## 3301	4426	42.386	31.771	43.614	11.84362	41.595	32.812	43.614
## 3302	2796	32.725	26.316	37.805	11.48909	30.079	29.825	36.585
## 3303	6481	34.023	27.797	37.879	10.08218	34.223	28.814	37.879
## 3304	20897	45.676	32.558	51.671	19.11295	44.523	31.190	49.125
## 3305	17181	29.230	22.581	34.400	11.81935	29.055	23.349	35.636
## 3306	5988	51.369	41.538	55.556	14.01709	47.929	43.077	51.184
## 3307	1001	25.974	22.388	28.750	6.36194	22.478	19.403	22.500
## 3308	9890	36.431	33.043	38.786	5.74280	35.571	33.043	37.291
## 3309	4030	37.395	32.941	34.722	1.78105	35.012	30.980	34.722
## 3310	7941	44.566	41.275	49.008	7.73333	43.420	38.926	47.450
## 3311	761	27.201	30.000	30.841	0.84112	25.099	10.000	28.037
## 3312	20713	35.948	29.249	40.999	11.75009	35.442	27.536	40.819
## 3313	1538	19.506	18.310	24.242	5.93257	20.026	15.493	22.727
## 3314	2933	41.016	32.979	44.608	11.62912	39.993	31.915	42.892
## 3315	3148	40.343	30.000	44.804	14.80370	40.248	30.000	43.649
## 3316	7344	38.875	31.081	42.744	11.66246	38.821	31.081	43.837
## 3317	1148	46.254	37.500	52.212	14.71239	45.557	34.375	52.212
## 3318	2034	80.728	82.022	77.551	-4.47145	79.302	77.528	78.571
## 3319	1010	52.970	48.000	68.224	20.22430	50.297	42.000	58.879
## 3320	8257	25.094	21.717	29.712	7.99437	25.869	21.212	30.962

## 3321	3435	29.491	25.444	29.655	4.21139	29.840	21.893	30.690
## 3322	2456	25.448	26.852	29.688	2.83565	26.099	25.000	36.328
## 3400	2435	41.848	32.743	51.643	18.89983	30.021	24.779	33.803
## 3401	55	54.545	80.000	33.333	-46.66667	38.182	40.000	33.333
## 3402	46	63.043	80.000	50.000	-30.00000	52.174	80.000	50.000
## 3403	276	56.884	9.091	67.857	58.76623	37.681	27.273	46.429
## 3404	143	51.049	0.000	50.000	50.00000	37.762	0.000	37.500
## 3500	1119	34.763	27.273	28.409	1.13636	24.397	27.273	22.727
## 3501	7	28.571	NaN	20.000	NaN	14.286	NaN	0.000
## 3502	0	NA	NA	NA	NA	NA	NA	NA
## 3503	2	0.000	NaN	NaN	NaN	50.000	NaN	NaN
## 3504	552	21.920	5.882	25.000	19.11765	14.130	5.882	10.714
## 3505	231	41.126	66.667	42.308	-24.35897	23.377	66.667	23.077
## 3506	100	27.000	11.111	33.333	22.22222	11.000	0.000	16.667
## 3600	800	49.250	44.828	52.174	7.34633	46.000	48.276	43.478
## 3601	493	49.493	66.667	50.746	-15.92040	48.276	66.667	44.776
## 3602	166	68.072	NaN	57.143	NaN	57.229	NaN	71.429
## 3603	121	33.058	21.429	30.000	8.57143	32.231	21.429	20.000
## 3604	17	17.647	NaN	NaN	NaN	23.529	NaN	NaN
## 3605	230	52.174	0.000	35.714	35.71429	51.304	0.000	42.857
## 3606	0	NA	NA	NA	NA	NA	NA	NA
## 3607	234	25.641	11.111	25.000	13.88889	19.231	33.333	25.000
## 3608	0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
## 3609	449	81.737	81.818	70.000	-11.81818	75.501	75.758	60.000
## 3610	351	34.473	11.765	48.000	36.23529	24.217	5.882	40.000
## 3611	385	58.961	82.353	58.065	-24.28843	40.519	35.294	38.710
## 3612	3655	44.542	55.063	37.346	-17.71761	38.523	54.430	34.877
## 3613	101	43.564	0.000	36.364	36.36364	14.851	0.000	9.091
## 3614	1152	34.115	35.714	39.623	3.90836	20.833	35.714	16.981
## 3615	0	NA	NA	NA	NA	NA	NA	NA
## 3616	1242	62.802	69.643	61.682	-7.96061	56.200	58.929	54.206
##	ChgN	1CtyMF96	1CtyMF14					
## 1000	9.39683	175	225					
## 1100	7.98441	227	947					
## 1101	10.04651	43	50					
## 1102	5.44872	120	117					
## 1103	8.40236	205	298					
## 1104	1.47626	179	207					
## 1105	3.15142	431	714					
## 1106	-0.77436	199	183					
## 1107	11.19895	44	69					
## 1108	18.87755	49	20					
## 1109	7.47995	103	115					
## 1110	4.14259	303	221					
## 1111	9.20304	51	62					
## 1200	8.77193	45	133					
## 1201	7.89069	344	757					
## 1202	13.36272	373	1229					
## 1203	10.61169	287	641					
## 1204	1.70886	80	237					

## 1205	32.92683	41	86
## 1206	48.27586	4	29
## 1207	12.87879	132	297
## 1208	11.04389	200	638
## 1209	70.00000	1	10
## 1210	9.64912	38	138
## 1211	13.02779	226	743
## 1212	3.51572	104	323
## 1213	11.74948	92	525
## 1300	12.84307	354	1098
## 1301	NaN	0	6
## 1302	2.36842	20	95
## 1303	3.71818	1369	1070
## 1304	4.97433	246	228
## 1305	9.57534	200	292
## 1306	9.62963	180	270
## 1307	6.71034	680	541
## 1308	5.43001	177	216
## 1309	3.36588	139	168
## 1310	11.16160	137	151
## 1311	1.45667	615	579
## 1312	6.36126	1007	847
## 1313	9.23811	169	258
## 1314	13.89728	331	331
## 1315	7.26821	151	120
## 1400	8.78929	234	254
## 1401	-0.10246	112	122
## 1402	14.10636	111	155
## 1403	4.35897	150	390
## 1404	16.74969	73	66
## 1405	9.49850	241	267
## 1406	3.71366	123	236
## 1407	17.56848	211	331
## 1408	5.61158	452	527
## 1409	11.39944	91	166
## 1410	-2.52167	54	47
## 1500	1.05893	134	253
## 1501	50.00000	6	6
## 1502	-1.19671	77	191
## 1503	2.43643	292	341
## 1504	-16.66667	2	3
## 1505	-5.52846	41	90
## 1506	0.00000	7	5
## 1507	-6.66667	15	15
## 1508	15.41667	30	32
## 1600	0.35449	859	821
## 1601	34.96241	14	19
## 1602	4.64110	247	233
## 1603	2.43697	70	119
## 1604	5.87454	256	206

## 1605	1.51207	617	519
## 1606	3.06278	573	513
## 1607	6.76461	149	164
## 1700	4.44101	91	121
## 1701	NaN	0	17
## 1702	0.12032	122	218
## 1703	7.46753	44	154
## 1704	15.90909	33	56
## 1705	12.72197	79	139
## 1706	0.51561	208	552
## 1707	20.58442	44	70
## 1708	0.00000	61	61
## 1709	-5.92992	49	106
## 1710	-4.15451	206	163
## 1711	13.84615	45	65
## 1712	-1.62050	183	261
## 1800	-7.12963	40	54
## 1801	NaN	0	3
## 1802	1.04628	35	71
## 1803	-3.49199	129	107
## 1804	-2.55620	69	127
## 1900	17.43802	121	110
## 1901	0.63757	87	128
## 1902	13.07125	99	185
## 1903	3.33333	10	45
## 1904	4.20644	163	223
## 1905	10.52632	16	19
## 1906	3.65027	122	150
## 1907	0.51865	140	157
## 1908	1.38889	108	136
## 1909	1.00334	13	46
## 1910	8.39260	114	148
## 1911	-5.54847	96	98
## 1912	2.81669	118	173
## 1913	13.40580	12	23
## 2000	7.98909	45	163
## 2001	-9.75728	20	103
## 2002	11.13256	551	893
## 2003	6.99144	144	266
## 2100	13.51496	26	144
## 2101	-81.25000	1	16
## 2102	-4.86111	54	112
## 2103	-4.39189	56	74
## 2104	-4.59459	20	37
## 2105	1.28353	49	159
## 2200	3.15341	330	192
## 2201	-1.03448	20	58
## 2202	-0.04605	43	101
## 2203	7.84314	4	51
## 2204	1.70503	69	238

## 2205	5.64034	83	270
## 2206	9.09091	15	11
## 2207	-0.69597	78	140
## 2208	1.92826	410	425
## 2209	14.05405	8	185
## 2210	2.72358	128	362
## 2211	6.20403	106	236
## 2212	4.28850	38	54
## 2213	3.97059	40	136
## 2214	2.38095	21	42
## 2215	5.97750	81	158
## 2216	16.66667	45	70
## 2300	10.47222	137	294
## 2301	6.79916	100	239
## 2302	-7.50605	14	59
## 2303	-0.56119	188	327
## 2304	2.78705	99	245
## 2305	0.82905	53	132
## 2306	-15.47170	25	106
## 2307	5.43053	84	73
## 2308	7.30608	200	526
## 2309	2.77072	57	145
## 2310	-2.09790	65	143
## 2311	0.61856	40	97
## 2312	9.50055	106	170
## 2400	6.45510	59	188
## 2401	NaN	0	3
## 2402	6.79794	119	67
## 2403	11.12156	339	386
## 2404	8.08156	242	244
## 2405	0.60471	73	111
## 2406	8.06154	81	134
## 2500	4.22644	182	393
## 2501	25.00000	8	8
## 2502	7.84314	36	153
## 2503	3.97952	255	203
## 2504	6.27326	137	355
## 2505	5.73467	528	344
## 2506	5.06281	284	185
## 2507	5.46603	112	82
## 2508	4.99263	265	256
## 2600	4.88606	185	255
## 2601	18.91892	13	37
## 2602	9.90260	42	88
## 2603	-9.37500	16	64
## 2604	3.46557	218	364
## 2605	4.42308	20	104
## 2606	5.10511	27	37
## 2607	4.76190	22	42
## 2608	-0.61633	22	59



## 2609	2.85714	15	35
## 2610	3.68387	78	79
## 2611	1.13157	88	237
## 2612	3.02198	13	28
## 2613	0.81340	169	355
## 2614	1.44231	64	117
## 2700	10.59236	705	2081
## 2701	8.35159	184	387
## 2702	-17.32543	44	69
## 2703	5.16154	54	235
## 2704	4.05797	30	46
## 2705	2.30091	299	582
## 2706	15.41463	75	205
## 2707	8.39002	18	49
## 2708	12.38028	71	100
## 2709	19.64286	8	7
## 2710	1.91816	23	34
## 2711	12.21418	41	128
## 2712	15.27557	130	261
## 2713	3.61052	59	192
## 2714	-3.39089	49	130
## 2715	11.63406	55	158
## 2716	6.17289	119	188
## 2717	12.80610	77	201
## 2718	-2.62085	17	101
## 2719	7.02933	190	506
## 2720	-0.65050	114	178
## 2721	15.55556	60	72
## 2722	-1.33576	27	61
## 2723	10.47078	176	224
## 2724	15.41682	82	134
## 2725	9.66134	130	318
## 2726	-0.39683	42	108
## 2727	-4.01003	42	133
## 2728	6.94269	235	555
## 2729	14.91517	238	263
## 2730	3.32088	185	369
## 2731	10.14177	161	262
## 2732	-4.30632	108	547
## 2733	10.48330	84	134
## 2734	23.19125	157	312
## 2735	-1.19292	221	462
## 2736	14.68023	129	336
## 2737	15.62092	136	225
## 2738	8.70836	493	1003
## 2739	7.43830	397	832
## 2740	5.09771	107	297
## 2741	4.13793	160	319
## 2742	-15.16324	83	193
## 2743	2.30769	95	104

## 2744	NaN	114	0
## 2745	11.06977	43	125
## 2746	2.72464	192	843
## 2747	-7.43746	51	87
## 2748	-16.93548	31	124
## 2800	6.94493	230	375
## 2801	20.83333	8	39
## 2802	5.52288	85	144
## 2803	13.35597	19	93
## 2804	13.01202	119	223
## 2805	1.13580	81	275
## 2806	4.37406	26	51
## 2807	36.36364	7	22
## 2808	6.40807	97	278
## 2809	14.06984	142	192
## 2900	4.91327	274	372
## 2901	NaN	0	15
## 2902	51.81554	34	162
## 2903	-14.28571	7	1
## 2904	NaN	0	15
## 2905	NaN	0	138
## 2906	-31.66667	24	30
## 2907	27.09360	7	29
## 2908	NaN	0	14
## 2909	41.11111	5	90
## 2910	10.11535	46	98
## 2911	2.15054	15	62
## 2912	-2.67559	13	23
## 2913	33.20388	5	103
## 2914	8.33333	2	96
## 2915	NaN	0	0
## 2916	-0.61915	157	179
## 2917	-36.36364	1	44
## 2918	NA	NA	NA
## 2919	10.43240	47	31
## 2920	-50.00000	2	1
## 2921	-4.72849	22	149
## 2922	-7.06897	29	20
## 2923	NaN	0	31
## 3000	23.33333	7	30
## 3001	17.50000	8	10
## 3002	6.38639	335	251
## 3003	3.83485	174	223
## 3004	11.46946	463	415
## 3005	15.68203	122	106
## 3100	-7.97954	20	391
## 3101	2.88915	64	106
## 3102	8.87994	66	79
## 3103	2.82416	99	147
## 3104	2.86485	326	548

## 3105	2.53968	70	99
## 3106	4.86666	89	166
## 3107	4.99861	108	266
## 3108	-7.90043	21	44
## 3109	-0.12045	81	123
## 3110	2.28031	62	87
## 3200	8.55148	282	382
## 3201	11.43695	33	62
## 3202	6.91176	176	340
## 3203	8.45856	215	511
## 3204	7.43085	302	544
## 3205	14.02859	225	443
## 3206	22.26647	49	123
## 3207	12.17986	152	411
## 3300	15.47468	145	397
## 3301	10.80196	192	415
## 3302	6.76080	114	246
## 3303	9.06523	295	462
## 3304	17.93452	731	1885
## 3305	12.28767	651	1375
## 3306	8.10705	195	549
## 3307	3.09701	67	80
## 3308	4.24764	345	1137
## 3309	3.74183	255	216
## 3310	8.52425	298	706
## 3311	18.03738	10	107
## 3312	13.28285	759	2222
## 3313	7.23431	71	132
## 3314	10.97726	94	408
## 3315	13.64896	70	433
## 3316	12.75590	222	1006
## 3317	17.83739	32	113
## 3318	1.04334	89	196
## 3319	16.87850	50	107
## 3320	9.74942	198	1040
## 3321	8.79616	169	290
## 3322	11.32812	108	256
## 3400	9.02406	113	213
## 3401	-6.66667	5	3
## 3402	-30.00000	5	2
## 3403	19.15584	11	28
## 3404	37.50000	2	24
## 3500	-4.54545	33	88
## 3501	NaN	0	5
## 3502	NA	NA	NA
## 3503	NaN	0	0
## 3504	4.83193	17	56
## 3505	-43.58974	3	26
## 3506	16.66667	9	6
## 3600	-4.79760	29	23

```
## 3601 -21.89055      12      67
## 3602      NaN       0       7
## 3603 -1.42857      14      10
## 3604      NaN       0       0
## 3605 42.85714       1      14
## 3606      NA      NA      NA
## 3607 -8.33333       9      20
## 3608      NaN       0       0
## 3609 -15.75758      33      20
## 3610 34.11765      17      25
## 3611  3.41556      17      31
## 3612 -19.55384     158     324
## 3613  9.09091       1      22
## 3614 -18.73315      42      53
## 3615      NA      NA      NA
## 3616 -4.72296      56     107
```

```
print("GenderTeamSize in 2018")
```

```
## [1] "GenderTeamSize in 2018"
```

```
print(GenderTeamSize)
```

```
##      Articles FirstF FirstM   FirstP LastF LastM   LastP
## 1000      582  4.185  3.611 7.931e-03 3.813 3.814 0.8188735
## 1100      661  4.116  3.408 3.469e-04 3.696 3.719 0.8306227
## 1101       56  2.871  2.275 1.431e-01 2.053 2.549 0.3140593
## 1102      139  3.883  3.258 1.781e-01 3.572 3.482 0.7148115
## 1103      293  2.959  2.398 1.752e-02 2.830 2.561 0.3099224
## 1104      228  3.379  3.363 9.135e-01 3.426 3.351 0.5757228
## 1105      612  3.353  2.942 4.788e-02 3.149 3.081 0.6474500
## 1106      197  2.843  2.918 8.313e-01 2.711 2.999 0.3316901
## 1107       88  3.152  3.217 8.893e-01 3.041 3.249 0.5520898
## 1108       13  2.914  1.861 3.263e-01 1.000 2.363 0.3345288
## 1109       84  3.947  3.126 1.341e-01 3.834 3.271 0.4624736
## 1110      201  3.492  2.646 1.901e-02 2.957 2.977 0.9514017
## 1111       71  4.855  3.221 5.462e-03 4.985 3.428 0.0443213
## 1200      111  1.198  1.297 4.121e-01 1.178 1.317 0.2270073
## 1201      749  1.752  1.575 2.160e-02 1.655 1.674 0.8803064
## 1202     1271  1.110  1.084 2.079e-01 1.096 1.092 0.7093820
## 1203      505  1.616  1.314 2.107e-05 1.562 1.374 0.0031312
## 1204      229  1.714  1.427 3.136e-03 1.490 1.551 0.7658854
## 1205       26  1.000  1.039 5.597e-01 1.080 1.000 0.1958301
## 1206       79  1.601  1.459 4.422e-01 1.569 1.488 0.6119891
## 1207      313  1.448  1.194 9.259e-04 1.421 1.210 0.0068680
## 1208      654  1.064  1.031 6.430e-02 1.061 1.034 0.0676790
## 1209       34  1.479  1.094 6.674e-02 1.233 1.241 0.8757289
## 1210      108  1.414  1.410 9.469e-01 1.337 1.452 0.5919941
## 1211      834  1.158  1.077 2.979e-03 1.134 1.088 0.0398462
## 1212      396  1.130  1.092 1.540e-01 1.146 1.085 0.1131984
## 1213      556  1.141  1.117 5.646e-01 1.179 1.071 0.0004111
```

## 1300	1019	4.498	3.918	2.055e-03	3.962	4.253	0.1606288
## 1301	30	3.340	3.888	1.000e+00	2.860	4.125	0.3245008
## 1302	59	4.463	4.701	5.949e-01	4.863	4.398	0.6349570
## 1303	611	4.832	4.258	1.607e-02	4.846	4.372	0.1281993
## 1304	245	4.549	3.220	4.888e-05	3.994	3.536	0.3351597
## 1305	282	3.844	3.450	1.186e-01	3.277	3.673	0.4045004
## 1306	268	6.131	6.333	7.692e-01	6.722	6.053	0.1028661
## 1307	407	5.155	4.127	6.016e-03	4.951	4.462	0.3712559
## 1308	126	5.232	4.488	1.648e-01	6.592	4.422	0.0046213
## 1309	146	4.337	3.466	6.866e-02	4.072	3.829	0.5971714
## 1310	140	5.181	4.481	2.455e-01	5.145	4.645	0.4564955
## 1311	470	5.192	4.092	1.798e-03	5.230	4.270	0.0061908
## 1312	680	4.761	4.266	9.778e-02	4.753	4.395	0.2020600
## 1313	227	5.240	5.693	3.654e-01	5.282	5.538	0.8750222
## 1314	333	4.188	3.703	2.291e-01	4.263	3.802	0.2571865
## 1315	92	4.228	3.041	2.860e-02	3.712	3.354	0.7063689
## 1400	276	1.710	1.669	8.213e-01	1.600	1.728	0.2863431
## 1401	123	1.948	1.724	2.196e-01	1.840	1.784	0.8604954
## 1402	177	1.663	1.633	7.247e-01	1.453	1.717	0.0695601
## 1403	388	1.833	1.688	2.074e-01	1.848	1.684	0.1667727
## 1404	58	2.544	2.002	9.837e-02	2.447	2.072	0.2921909
## 1405	306	2.010	1.871	2.431e-01	1.888	1.938	0.6877171
## 1406	186	2.024	2.058	5.655e-01	2.143	1.987	0.4699005
## 1407	227	1.805	1.792	8.596e-01	1.778	1.811	0.8830840
## 1408	568	1.889	1.982	2.726e-01	1.827	2.007	0.0747896
## 1409	215	1.761	1.668	4.194e-01	1.588	1.757	0.1609685
## 1410	30	1.644	1.443	4.243e-01	1.644	1.443	0.4243423
## 1500	269	4.195	3.220	2.758e-04	3.638	3.435	0.5191609
## 1501	28	2.405	3.275	3.940e-01	1.565	3.299	0.1082578
## 1502	198	3.881	3.893	8.684e-01	3.825	3.904	0.9793091
## 1503	377	5.081	4.518	6.614e-02	4.978	4.621	0.3422049
## 1504	NA	NA	NA	NA	NA	NA	NA
## 1505	75	6.132	5.342	3.060e-01	5.716	5.501	0.7573722
## 1506	10	6.769	3.659	2.301e-02	5.958	3.984	0.1593410
## 1507	78	3.269	2.802	4.692e-01	3.219	2.864	0.5500456
## 1508	89	3.390	3.283	9.045e-01	4.086	3.239	0.2018700
## 1600	855	5.119	4.178	3.958e-05	4.957	4.317	0.0111473
## 1601	49	3.547	2.335	1.255e-02	2.310	3.061	0.2468797
## 1602	160	3.928	3.468	1.402e-01	3.235	3.706	0.3749890
## 1603	66	4.541	3.442	5.231e-02	4.103	3.737	0.9350452
## 1604	179	4.164	4.027	9.713e-01	4.072	4.070	0.8528537
## 1605	316	4.604	4.326	2.536e-01	4.963	4.347	0.2638700
## 1606	475	4.030	3.394	1.746e-02	4.097	3.460	0.0408578
## 1607	144	4.264	3.647	1.103e-01	3.798	3.893	0.8148104
## 1700	161	2.306	2.288	9.453e-01	2.281	2.293	0.8052483
## 1701	27	1.799	1.787	1.000e+00	1.648	1.835	0.5758044
## 1702	42	2.543	2.096	3.689e-01	2.632	2.115	0.3556938
## 1703	112	2.266	2.871	9.903e-02	2.309	2.878	0.1395082
## 1704	58	2.564	2.961	4.860e-01	1.856	3.148	0.0212062
## 1705	95	2.318	2.418	7.074e-01	2.089	2.456	0.2046809

## 1706	317	2.161	2.517	1.035e-01	2.108	2.543	0.0405843
## 1707	63	3.452	2.841	6.701e-01	2.647	3.013	0.5133685
## 1708	55	3.443	2.467	8.899e-02	3.086	2.587	0.6288116
## 1709	126	2.342	2.280	8.356e-01	2.083	2.407	0.2086600
## 1710	188	2.526	2.091	1.726e-01	2.007	2.235	0.2754185
## 1711	75	2.431	2.808	3.204e-01	3.113	2.720	0.4206566
## 1712	220	2.939	2.599	2.752e-01	2.872	2.626	0.4880511
## 1800	74	2.426	2.039	2.088e-01	2.281	2.119	0.5850641
## 1801	NA	NA	NA	NA	NA	NA	NA
## 1802	99	1.777	2.020	3.672e-01	2.134	1.913	0.4059534
## 1803	153	1.916	2.285	1.134e-01	2.108	2.228	0.6469448
## 1804	158	2.104	1.832	4.228e-01	1.803	1.904	0.6130061
## 1900	150	2.397	2.790	3.862e-01	2.276	2.803	0.2531048
## 1901	95	2.733	2.805	9.731e-01	2.629	2.834	0.7269953
## 1902	219	3.091	3.097	8.551e-01	2.820	3.154	0.4627373
## 1903	33	3.157	2.334	2.016e-01	2.711	2.591	0.9328323
## 1904	164	2.120	2.070	7.061e-01	1.690	2.265	0.0159382
## 1905	28	1.585	2.199	4.278e-01	1.431	2.249	0.2160688
## 1906	115	3.213	2.812	2.178e-01	2.615	2.970	0.4364524
## 1907	213	3.386	2.703	3.124e-02	3.297	2.758	0.0928614
## 1908	106	3.232	2.790	3.355e-01	2.948	2.875	0.8181545
## 1909	110	2.230	2.354	6.373e-01	2.917	2.213	0.0226485
## 1910	101	3.228	2.709	1.981e-01	3.158	2.788	0.4666310
## 1911	107	2.921	2.385	8.855e-02	2.175	2.541	0.3793486
## 1912	137	2.989	2.608	2.257e-01	2.608	2.683	0.8528943
## 1913	36	4.141	2.863	9.942e-02	3.866	3.000	0.4147296
## 2000	119	1.457	1.436	8.403e-01	1.519	1.412	0.2928889
## 2001	179	1.473	1.310	1.731e-01	1.322	1.391	0.2592109
## 2002	807	1.630	1.530	1.142e-01	1.532	1.569	0.6056970
## 2003	269	1.684	1.640	7.247e-01	1.595	1.662	0.5958491
## 2100	152	3.245	2.836	4.152e-01	3.306	2.858	0.2801923
## 2101	36	3.224	2.412	1.993e-01	2.759	2.602	0.8604735
## 2102	202	2.809	3.230	1.280e-01	2.101	3.313	0.0001997
## 2103	134	2.833	2.890	6.772e-01	2.093	3.069	0.0060094
## 2104	93	2.844	3.218	2.123e-01	2.210	3.365	0.0045269
## 2105	355	2.899	3.201	1.789e-01	2.610	3.249	0.0094612
## 2200	160	2.930	2.407	2.030e-01	2.396	2.528	0.7462925
## 2201	78	2.321	2.294	8.435e-01	1.947	2.378	0.4821759
## 2202	94	2.995	2.654	5.260e-01	2.479	2.729	0.6902420
## 2203	63	2.563	2.607	9.043e-01	2.148	2.722	0.2374618
## 2204	305	4.154	3.718	1.255e-01	3.858	3.849	0.8736173
## 2205	360	2.806	2.580	2.070e-01	2.422	2.674	0.1291452
## 2206	24	2.828	2.365	6.657e-01	2.828	2.365	0.6657311
## 2207	66	2.855	2.582	5.011e-01	2.141	2.692	0.5277173
## 2208	185	3.665	3.132	2.816e-01	4.321	3.139	0.0494567
## 2209	269	2.632	3.028	1.153e-01	2.548	3.035	0.0609933
## 2210	552	3.606	3.039	7.267e-03	3.306	3.108	0.7176921
## 2211	310	3.628	2.899	9.221e-03	3.264	2.976	0.3349441
## 2212	54	2.747	3.229	6.914e-01	2.702	3.251	0.4928619
## 2213	174	2.708	2.415	2.994e-01	2.440	2.523	0.8104250

## 2214	44	1.494	2.448	8.026e-03	1.419	2.497	0.0019567
## 2215	242	3.264	2.708	9.744e-03	2.525	2.876	0.0430097
## 2216	94	1.356	1.485	4.292e-01	1.258	1.547	0.1090187
## 2300	298	2.573	2.557	9.795e-01	2.173	2.701	0.0181664
## 2301	175	1.499	1.902	9.889e-03	1.559	1.886	0.0463490
## 2302	44	3.600	2.927	3.239e-01	4.494	2.879	0.0109817
## 2303	360	3.254	3.096	7.277e-01	3.169	3.153	0.7665639
## 2304	211	4.224	3.877	2.904e-01	3.711	4.115	0.3452932
## 2305	153	3.608	3.327	3.163e-01	3.265	3.459	0.7918092
## 2306	161	2.233	2.392	6.439e-01	1.779	2.561	0.0049213
## 2307	134	4.224	3.634	2.139e-01	4.152	3.756	0.3369651
## 2308	619	2.091	2.005	4.122e-01	1.919	2.100	0.1184014
## 2309	255	2.838	2.590	3.026e-01	2.841	2.610	0.3195994
## 2310	203	4.082	3.698	1.784e-01	3.959	3.792	0.4679042
## 2311	135	3.999	3.569	3.603e-01	3.649	3.745	0.7798273
## 2312	183	3.131	3.255	9.803e-01	3.156	3.235	0.8289954
## 2400	220	3.740	3.351	3.097e-01	3.174	3.642	0.1765614
## 2401	8	6.836	6.160	5.590e-01	6.481	6.492	1.0000000
## 2402	68	3.730	3.863	9.206e-01	3.041	4.098	0.2085559
## 2403	323	5.722	4.232	1.084e-02	5.640	4.579	0.0249681
## 2404	224	4.416	4.336	8.543e-01	4.123	4.513	0.6260488
## 2405	103	5.464	4.445	3.668e-01	5.664	4.657	0.2838854
## 2406	101	5.920	6.050	8.286e-01	7.092	5.471	0.0559516
## 2500	630	4.150	3.580	1.835e-02	3.930	3.661	0.2801230
## 2501	26	1.737	2.712	1.657e-01	1.692	2.835	0.0928106
## 2502	137	4.785	3.576	6.535e-03	4.334	3.882	0.6050474
## 2503	159	4.547	3.690	7.006e-02	4.229	3.863	0.7321019
## 2504	235	4.500	3.711	4.793e-02	3.791	3.849	0.7222171
## 2505	142	4.613	3.673	3.634e-02	4.541	3.822	0.2641257
## 2506	143	4.571	3.703	1.023e-01	4.508	3.831	0.3963834
## 2507	94	3.810	3.686	6.193e-01	3.113	3.866	0.1499843
## 2508	224	4.659	3.823	1.745e-02	4.494	3.913	0.3599888
## 2600	256	1.720	1.473	9.192e-02	1.682	1.479	0.1002191
## 2601	27	1.516	1.231	1.852e-01	1.260	1.281	1.0000000
## 2602	93	1.704	1.225	2.357e-03	1.514	1.285	0.2233978
## 2603	71	1.452	1.481	9.083e-01	1.718	1.422	0.1479590
## 2604	372	2.297	1.896	1.136e-02	2.096	1.941	0.4886895
## 2605	117	2.656	2.352	5.527e-01	2.830	2.289	0.2149150
## 2606	67	3.058	2.284	5.148e-02	1.956	2.468	0.2272386
## 2607	46	2.119	1.530	6.990e-02	2.449	1.577	0.1512360
## 2608	55	1.682	1.356	1.172e-01	1.516	1.388	0.5700465
## 2609	23	1.189	1.076	4.873e-01	1.189	1.076	0.4873119
## 2610	105	1.811	1.594	4.240e-01	2.169	1.584	0.1494760
## 2611	269	2.867	2.485	6.938e-02	2.628	2.542	0.7743745
## 2612	21	1.414	2.022	1.986e-01	1.260	2.020	0.1647401
## 2613	382	2.414	2.069	6.242e-02	2.369	2.081	0.1300047
## 2614	130	1.934	1.995	6.303e-01	2.077	1.973	0.9913155
## 2700	1288	3.138	2.934	1.140e-01	2.834	3.165	0.0470408
## 2701	497	4.802	3.686	1.402e-03	4.405	4.216	0.7158680
## 2702	53	4.188	4.788	5.824e-01	3.882	4.711	0.2828816

## 2703	114	4.463	2.796	1.921e-04	3.897	3.689	0.8212033
## 2704	32	6.972	3.972	3.865e-02	7.331	4.667	0.1389113
## 2705	511	5.467	4.782	3.206e-01	6.017	4.783	0.0106236
## 2706	154	3.783	3.764	9.506e-01	3.902	3.728	0.6158531
## 2707	37	3.488	2.486	1.423e-01	3.375	2.549	0.2814558
## 2708	116	3.510	3.700	5.896e-01	3.515	3.681	0.6977570
## 2709	18	1.919	1.316	2.616e-01	1.974	1.320	0.1927027
## 2710	27	3.220	1.771	4.915e-02	2.836	2.197	0.3404147
## 2711	101	3.171	3.264	7.791e-01	3.401	3.207	0.8659668
## 2712	229	5.069	4.666	3.336e-01	4.984	4.802	0.6967708
## 2713	203	5.079	4.384	3.613e-01	5.258	4.391	0.1617946
## 2714	91	4.007	3.554	6.637e-01	3.954	3.784	0.5002279
## 2715	151	5.159	5.566	2.176e-01	5.611	5.401	0.9118878
## 2716	174	5.285	3.795	1.341e-02	5.181	4.203	0.1234740
## 2717	170	4.117	4.396	4.687e-01	4.058	4.388	0.2471861
## 2718	147	4.373	4.219	8.799e-01	4.261	4.303	0.9627244
## 2719	658	3.092	2.815	1.478e-01	2.936	2.969	0.9940056
## 2720	122	5.011	4.897	8.995e-01	3.866	5.489	0.0072809
## 2721	71	5.908	5.415	6.706e-01	5.405	5.629	0.8292546
## 2722	45	4.518	4.550	9.908e-01	3.957	4.628	0.4795567
## 2723	236	5.601	4.639	2.976e-01	5.464	4.908	0.2517720
## 2724	117	5.725	4.761	2.372e-01	5.122	5.158	0.9204065
## 2725	348	4.838	4.543	4.769e-01	4.550	4.794	0.6490385
## 2726	144	5.184	4.762	9.567e-01	5.972	4.503	0.0583433
## 2727	83	4.581	3.743	4.499e-01	4.134	3.973	0.8016339
## 2728	538	4.884	4.432	2.243e-01	4.595	4.637	0.9977020
## 2729	262	3.439	3.632	2.552e-01	3.213	3.867	0.0104223
## 2730	386	5.695	5.966	3.457e-01	5.921	5.800	0.7671678
## 2731	251	4.062	3.680	3.771e-01	4.121	3.746	0.3190890
## 2732	673	3.653	3.499	3.080e-01	3.179	3.638	0.0173462
## 2733	88	3.622	3.814	5.752e-01	3.330	3.892	0.3365357
## 2734	284	2.729	2.693	6.331e-01	2.409	2.963	0.0204476
## 2735	444	4.154	3.457	1.341e-02	3.856	3.870	0.8708856
## 2736	353	5.499	4.703	8.581e-02	5.295	4.993	0.4648491
## 2737	236	4.511	4.366	7.307e-01	4.234	4.475	0.6355113
## 2738	995	3.426	3.387	9.847e-01	3.203	3.606	0.0152246
## 2739	970	3.551	3.323	2.287e-01	3.461	3.450	0.7396427
## 2740	265	5.359	4.899	3.303e-01	5.171	5.067	0.8448755
## 2741	380	5.109	5.473	4.125e-01	5.047	5.432	0.4783100
## 2742	202	3.474	2.967	8.259e-02	3.415	3.096	0.4741203
## 2743	90	3.076	3.241	5.818e-01	2.863	3.411	0.1972776
## 2744	NA	NA	NA	NA	NA	NA	NA
## 2745	132	4.867	4.956	7.940e-01	4.861	4.938	0.7510483
## 2746	639	4.321	4.161	7.034e-01	3.842	4.267	0.0759628
## 2747	59	5.725	4.069	2.058e-01	5.217	4.714	0.9321050
## 2748	134	5.119	4.485	4.373e-01	5.380	4.532	0.5539963
## 2800	345	4.055	3.601	1.109e-01	3.266	3.977	0.0025926
## 2801	30	4.792	4.009	4.054e-01	4.631	4.193	0.5389008
## 2802	149	3.419	3.329	8.673e-01	3.603	3.217	0.4266239
## 2803	119	4.932	4.572	9.074e-01	4.793	4.768	0.9303270



## 2804	201	5.393	4.086	2.491e-03	5.281	4.321	0.0183693
## 2805	308	3.400	2.905	3.873e-02	3.159	3.105	0.9570112
## 2806	35	4.291	3.751	5.305e-01	4.436	3.534	0.3290045
## 2807	22	3.615	4.826	2.161e-01	4.748	4.011	0.4528877
## 2808	271	4.609	4.443	9.495e-01	4.280	4.660	0.3859759
## 2809	140	4.435	4.124	5.755e-01	3.964	4.337	0.2710532
## 2900	280	2.347	1.979	5.954e-02	2.273	2.133	0.4265074
## 2901	34	2.889	3.013	9.444e-01	2.851	3.053	0.9028491
## 2902	121	2.553	3.402	5.662e-02	2.312	3.509	0.0060478
## 2903	4	2.714	4.000	1.000e+00	2.520	5.000	0.3457786
## 2904	11	2.068	1.495	5.319e-01	1.719	1.944	0.8404757
## 2905	157	1.525	1.353	1.200e-01	1.463	1.504	0.8383472
## 2906	30	3.543	3.184	5.910e-01	3.452	3.278	0.7209577
## 2907	19	2.847	2.586	1.000e+00	3.203	2.486	0.4186653
## 2908	20	2.282	3.448	1.010e-01	2.293	3.221	0.1349312
## 2909	68	2.960	2.498	4.800e-01	2.689	3.155	0.1784591
## 2910	95	1.882	1.392	1.242e-02	1.856	1.435	0.0777198
## 2911	41	1.757	1.803	8.893e-01	2.128	1.619	0.2141224
## 2912	14	3.185	2.000	1.873e-01	3.018	1.000	0.1575442
## 2913	97	2.268	2.026	4.984e-01	2.173	2.564	0.1211264
## 2914	64	2.554	2.882	7.793e-01	2.475	3.315	0.1011066
## 2915	NA	NA	NA	NA	NA	NA	NA
## 2916	211	4.291	4.656	7.183e-02	4.374	4.441	0.7256643
## 2917	46	3.118	2.702	4.642e-01	2.829	3.211	0.6943576
## 2918	NA	NA	NA	NA	NA	NA	NA
## 2919	42	2.374	2.484	9.200e-01	2.175	3.009	0.0684006
## 2920	NA	NA	NA	NA	NA	NA	NA
## 2921	148	2.556	2.767	7.203e-01	2.342	2.950	0.0722523
## 2922	20	2.388	1.576	2.001e-01	2.297	1.586	0.2904089
## 2923	24	2.705	2.282	4.563e-01	2.699	2.342	0.5658731
## 3000	45	4.576	2.943	6.332e-02	3.737	3.445	0.8200111
## 3001	10	5.477	3.145	5.074e-01	5.958	2.472	0.1939309
## 3002	193	4.886	4.665	6.941e-01	5.037	4.692	0.6983020
## 3003	215	3.919	3.836	8.780e-01	3.735	3.922	0.4557848
## 3004	276	4.726	4.192	3.412e-01	4.831	4.301	0.1055735
## 3005	109	5.067	4.706	8.355e-01	4.668	4.945	0.3435542
## 3100	514	4.479	2.789	4.786e-09	3.963	2.940	0.0003134
## 3101	160	2.669	2.154	1.959e-01	2.609	2.166	0.2991478
## 3102	76	2.836	2.746	9.585e-01	2.810	2.757	0.9879663
## 3103	189	2.681	2.135	2.572e-02	2.510	2.183	0.1884529
## 3104	602	3.695	2.982	4.866e-04	3.558	3.060	0.0723229
## 3105	187	3.963	3.202	5.547e-02	3.595	3.293	0.4674799
## 3106	110	3.326	2.102	9.454e-03	2.069	2.308	0.6653364
## 3107	352	3.752	3.442	5.469e-01	3.538	3.482	0.7124839
## 3108	34	5.510	3.580	1.466e-01	7.573	3.571	0.0437966
## 3109	134	2.022	1.884	6.615e-01	2.405	1.859	0.1003778
## 3110	62	4.065	3.221	1.192e-01	3.386	3.470	0.5720992
## 3200	384	2.469	2.334	4.994e-01	2.441	2.374	0.5907588
## 3201	42	2.847	1.608	7.035e-04	2.555	1.945	0.1470176
## 3202	411	2.702	2.288	8.774e-03	2.589	2.444	0.4477097

## 3203	504	2.784	2.196	4.006e-04	2.597	2.538	0.6395457
## 3204	563	2.770	2.530	9.399e-02	2.628	2.765	0.3988191
## 3205	472	2.837	2.476	1.073e-02	2.703	2.625	0.5353767
## 3206	138	3.066	2.857	6.785e-01	3.001	2.949	0.7569280
## 3207	459	2.151	1.696	3.508e-05	2.015	1.812	0.0386225
## 3300	380	1.560	1.442	1.772e-01	1.553	1.449	0.4154657
## 3301	442	1.960	1.574	7.275e-05	1.905	1.631	0.0054311
## 3302	233	1.599	1.389	2.574e-02	1.478	1.452	0.5976993
## 3303	440	1.436	1.265	8.528e-03	1.342	1.328	0.9612938
## 3304	1657	1.842	1.603	7.939e-07	1.739	1.728	0.7161062
## 3305	1226	1.552	1.455	4.058e-02	1.502	1.485	0.8320675
## 3306	623	2.125	2.058	4.835e-01	2.078	2.127	0.7466206
## 3307	78	2.837	2.491	5.207e-01	2.710	2.648	0.9189876
## 3308	1168	1.452	1.358	2.084e-02	1.484	1.333	0.0001614
## 3309	191	1.774	1.825	6.298e-01	1.835	1.782	0.4974910
## 3310	595	1.558	1.318	1.437e-04	1.496	1.391	0.0574569
## 3311	136	1.915	1.909	9.732e-01	1.957	1.884	0.7391316
## 3312	2304	1.462	1.299	1.439e-07	1.425	1.327	0.0047699
## 3313	149	2.620	2.298	2.298e-01	2.461	2.377	0.6166264
## 3314	507	1.277	1.124	8.775e-04	1.233	1.159	0.0856633
## 3315	502	1.331	1.320	6.801e-01	1.366	1.285	0.0661482
## 3316	1255	1.145	1.130	3.833e-01	1.147	1.128	0.1186004
## 3317	108	1.293	1.265	7.920e-01	1.240	1.341	0.2713358
## 3318	241	1.335	1.188	1.200e-01	1.306	1.262	0.6463846
## 3319	124	2.001	1.805	3.543e-01	1.919	1.924	1.0000000
## 3320	943	1.190	1.186	9.466e-01	1.211	1.176	0.4971350
## 3321	228	1.924	1.504	9.693e-04	1.851	1.563	0.0207610
## 3322	207	1.621	1.451	1.749e-01	1.693	1.402	0.0072342
## 3400	266	3.417	3.101	1.427e-01	3.260	3.315	0.8058589
## 3401	10	2.962	2.289	4.847e-01	2.340	3.047	0.5859730
## 3402	10	2.416	6.649	2.064e-02	3.457	2.000	0.4795001
## 3403	34	4.136	3.236	4.545e-01	4.067	3.642	0.8067023
## 3404	14	4.169	3.982	1.000e+00	3.471	4.409	0.2436898
## 3500	92	2.958	2.115	1.292e-02	2.476	2.473	0.9714405
## 3501	4	6.000	2.714	3.458e-01	2.000	3.915	0.6373519
## 3502	NA	NA	NA	NA	NA	NA	NA
## 3503	NA	NA	NA	NA	NA	NA	NA
## 3504	32	2.958	3.575	3.572e-01	1.682	3.646	0.0447210
## 3505	15	2.723	2.520	8.207e-01	2.870	2.479	0.7619868
## 3506	5	1.000	4.527	4.000e-01	1.732	5.192	0.2000000
## 3600	25	1.456	2.683	3.297e-02	1.664	2.423	0.1012140
## 3601	42	1.868	1.644	5.264e-01	1.709	1.824	0.6886857
## 3602	15	3.956	2.901	2.848e-01	3.711	3.120	0.5927843
## 3603	8	3.000	3.360	3.902e-01	4.481	2.702	0.2187982
## 3604	6	2.862	10.000	2.134e-01	2.000	4.681	0.4606335
## 3605	42	2.544	2.626	7.396e-01	2.721	2.536	0.9360949
## 3606	NA	NA	NA	NA	NA	NA	NA
## 3607	16	3.557	1.861	8.130e-02	4.932	1.814	0.0129022
## 3608	4	1.442	1.000	1.000e+00	1.442	1.000	1.0000000
## 3609	13	3.502	1.000	3.455e-02	3.359	2.268	0.4554137

```
## 3610      20  3.399  2.399 2.605e-01 4.300 2.434 0.0821361
## 3611      43  2.813  3.653 2.257e-01 3.002 3.437 0.4046176
## 3612     417  3.437  3.128 2.201e-01 3.150 3.283 0.2881533
## 3613       5  3.464  3.557 1.000e+00    NA    NA        NA
## 3614      71  4.338  5.841 2.275e-02 5.361 5.187 0.5541818
## 3615     NA    NA    NA    NA    NA    NA        NA
## 3616      83  3.060  2.771 2.431e-01 2.757 3.217 0.2370328
```

```
print(RegCoef)
```

```
##          FFA1          FLA1          2          3          4          5+
## 1000  5.046e-03 -4.442e-02  1.578010  1.644e+00  1.7111559  1.783314
## 1100 -1.453e-02 -8.176e-03  0.354693  4.498e-01  0.4744013  0.540996
## 1101 -4.529e-02  1.771e-03  0.235198  2.359e-01  0.2576939  0.368789
## 1102  5.266e-02  2.603e-02  0.214842  3.549e-01  0.4190961  0.495236
## 1103  7.376e-02  6.322e-02  0.249954  2.905e-01  0.3442531  0.396365
## 1104 -8.001e-03 -3.115e-02  0.129345  1.646e-01  0.1988604  0.277557
## 1105 -1.758e-02  1.144e-02  0.188289  2.224e-01  0.2513098  0.325033
## 1106  4.401e-02 -9.013e-02  0.438648  5.393e-01  0.5930421  0.675836
## 1107 -2.357e-02 -4.233e-02  0.195411  1.099e-01  0.1519240  0.229877
## 1108  3.866e-02 -1.341e-02  0.282700  3.391e-01  0.3995438  0.497869
## 1109 -2.331e-02 -1.513e-02  0.343342  4.053e-01  0.4341208  0.557662
## 1110  1.985e-02 -3.806e-02  0.351165  4.367e-01  0.5248356  0.611577
## 1111 -3.927e-02 -4.403e-02  0.061829  3.802e-02  0.0859861  0.139393
## 1200  1.242e-01 -1.108e-01  0.495112  8.329e-01  0.9449154  1.223570
## 1201  1.451e-02  6.032e-04  0.162653  1.781e-01  0.1849234  0.250630
## 1202 -2.696e-02  1.016e-02  0.286219  3.060e-01  0.6567512  0.227601
## 1203  3.003e-02  2.179e-02  0.375682  4.602e-01  0.5091903  0.586602
## 1204  3.147e-03 -1.266e-01  0.310335  4.264e-01  0.4590101  0.406881
## 1205  1.488e-01 -1.311e-01  0.189221 -6.144e-01  0.0440973  1.107972
## 1206  9.479e-02 -8.559e-02  0.269899  8.008e-01  0.5207742  0.244596
## 1207  1.555e-01  4.925e-02  0.385891  5.588e-01  0.4696433  0.581610
## 1208  8.380e-02 -5.985e-02  0.360166  6.512e-01  0.8145663 -0.036155
## 1209  4.211e-02 -8.553e-02  0.314523  4.599e-01  0.5285659  0.461538
## 1210  1.116e-01  6.093e-02  0.459261  3.962e-01  0.6870033  0.369323
## 1211 -7.546e-03 -6.993e-03  0.198870  2.533e-01  0.0756230  0.034913
## 1212  1.158e-01 -2.029e-02  0.429180  4.583e-01  0.6932570  0.403795
## 1213  8.805e-02  1.821e-02  0.158874  1.868e-01  0.4272433  0.069686
## 1300  4.372e-03 -1.571e-02  0.282903  3.043e-01  0.3315946  0.446310
## 1301     NA     NA     NA     NA     NA     NA
## 1302  2.768e-02  7.426e-02  0.091318  1.985e-01  0.1998061  0.279667
## 1303  3.313e-03 -8.800e-03  0.155383  1.629e-01  0.1887309  0.259794
## 1304 -3.801e-02 -4.409e-03  0.172719  1.665e-01  0.1783435  0.243472
## 1305 -2.343e-02 -3.805e-02  0.362121  4.035e-01  0.4442655  0.537574
## 1306 -4.196e-02 -3.481e-02  0.581406  6.094e-01  0.6077648  0.781660
## 1307 -3.592e-03  4.792e-03  0.181627  2.092e-01  0.2235775  0.302503
## 1308 -7.017e-02 -1.218e-01  0.246435  4.370e-01  0.5471120  0.673980
## 1309  3.413e-03  3.586e-03  0.113609  1.060e-01  0.1282551  0.167476
## 1310  8.257e-04  9.720e-03  0.293744  2.961e-01  0.3854027  0.465436
## 1311 -2.608e-02 -1.524e-02  0.285745  3.206e-01  0.3476799  0.417975
```

##	1312	-5.697e-03	-4.883e-03	0.250166	2.650e-01	0.2948604	0.356099
##	1313	-9.721e-03	-3.131e-02	0.274346	4.665e-01	0.4474986	0.556547
##	1314	-2.161e-02	-1.544e-02	0.074861	1.618e-01	0.1677693	0.254090
##	1315	-4.941e-02	2.430e-02	-0.006776	2.364e-02	0.0363357	0.093254
##	1400	-4.217e-02	6.280e-03	0.205918	3.230e-01	0.3248171	0.187501
##	1401	1.482e-02	2.075e-02	0.214660	2.251e-01	0.3170697	0.314624
##	1402	2.876e-02	5.807e-02	0.061751	8.355e-02	0.1678780	0.429603
##	1403	-3.401e-02	-2.637e-03	0.188286	3.402e-01	0.2870465	0.064329
##	1404	7.820e-02	-1.244e-02	0.158346	1.905e-01	0.2954716	0.099179
##	1405	-1.531e-03	8.789e-03	0.180188	2.526e-01	0.2559129	0.183554
##	1406	5.449e-02	1.752e-02	0.217023	2.681e-01	0.2912957	0.167820
##	1407	4.796e-02	2.415e-02	0.230894	2.242e-01	0.3317392	0.404933
##	1408	-1.467e-02	-1.610e-02	0.208141	2.549e-01	0.2692112	0.247443
##	1409	3.275e-02	-9.763e-03	0.046971	5.395e-02	-0.0841455	0.135640
##	1410	5.276e-02	-5.922e-02	0.255949	2.994e-01	0.2966955	0.516610
##	1500	-5.177e-02	-6.683e-02	0.441766	4.552e-01	0.5177913	0.555630
##	1501	-4.359e-01	-2.168e-01	0.100135	8.524e-01	1.0385341	0.968317
##	1502	-2.934e-02	1.279e-02	0.386112	4.235e-01	0.4795934	0.580228
##	1503	-3.291e-02	-1.224e-02	0.061054	5.997e-02	0.1266987	0.124399
##	1504	6.979e-02	-4.122e-01	0.495019	2.163e-01	-0.0627441	0.253416
##	1505	-4.316e-02	-6.571e-03	0.072038	1.205e-01	0.1709798	0.193704
##	1506	1.010e-01	1.375e-01	0.336899	5.355e-01	0.3064229	0.307936
##	1507	-4.876e-02	5.772e-02	0.098549	2.057e-01	0.4386170	0.386079
##	1508	3.945e-02	-6.897e-02	0.632749	8.130e-01	0.8845397	0.844312
##	1600	-1.061e-02	2.812e-03	0.342845	3.593e-01	0.3959724	0.434080
##	1601	9.670e-02	2.034e-02	0.483812	4.344e-01	0.3226468	0.500830
##	1602	3.714e-02	1.499e-02	0.281857	3.060e-01	0.3594230	0.409281
##	1603	-2.491e-02	-2.230e-02	0.130671	9.704e-02	0.1502416	0.148148
##	1604	1.602e-02	-2.679e-02	0.032289	4.039e-02	0.0870661	0.126692
##	1605	-2.298e-03	-3.101e-02	0.037794	6.462e-02	0.0987013	0.136439
##	1606	1.228e-02	-4.368e-02	0.110985	8.743e-02	0.1177447	0.137476
##	1607	4.538e-02	-3.109e-03	0.199052	2.369e-01	0.2742088	0.271410
##	1700	8.549e-02	4.630e-02	0.230635	2.538e-01	0.4332403	0.524480
##	1701	-1.697e-01	3.273e-01	0.301274	2.103e-01	0.1291827	-0.176375
##	1702	1.967e-02	8.584e-02	0.181051	1.975e-01	0.3295978	0.270383
##	1703	-5.041e-02	-5.745e-02	0.281818	3.554e-01	0.3897963	0.598602
##	1704	-1.376e-01	6.276e-03	0.300937	4.583e-01	0.4772799	0.598826
##	1705	1.527e-02	-5.428e-02	0.217157	3.613e-01	0.4832905	0.406383
##	1706	-2.331e-02	-6.322e-03	0.217086	2.500e-01	0.3314119	0.338358
##	1707	-3.641e-02	6.474e-02	0.113816	1.516e-01	0.2187268	0.333783
##	1708	1.693e-02	-1.397e-03	0.403870	3.845e-01	0.5375446	0.502927
##	1709	2.230e-02	2.164e-02	0.299776	4.502e-01	0.4070186	0.402095
##	1710	1.077e-04	-3.213e-03	0.166729	3.065e-01	0.3385862	0.294406
##	1711	-2.573e-02	3.976e-02	0.165659	1.739e-01	0.2391695	0.394625
##	1712	5.233e-02	2.292e-02	0.267664	3.137e-01	0.4015466	0.428335
##	1800	2.456e-02	7.830e-02	0.079928	1.926e-01	0.0721883	0.314070
##	1801	NA	NA	NA	NA	NA	NA
##	1802	1.140e-01	5.442e-03	-0.076275	-5.078e-05	0.1116421	-0.251369
##	1803	4.554e-02	3.414e-02	0.134295	2.123e-01	0.3031274	0.111955
##	1804	-6.539e-02	6.955e-02	0.198657	3.119e-01	0.3867608	0.631975

## 1900	-8.010e-02	-4.459e-02	0.272184	3.942e-01	0.3716702	0.692247
## 1901	-3.687e-02	-9.454e-03	0.228139	2.653e-01	0.2942800	0.344617
## 1902	-5.500e-03	7.656e-03	0.411457	4.662e-01	0.5226573	0.602920
## 1903	-7.640e-02	6.664e-02	0.099979	3.090e-01	0.2045904	0.499746
## 1904	1.366e-02	3.183e-02	0.181280	1.952e-01	0.1882081	0.290771
## 1905	-9.190e-02	-2.713e-01	0.170682	4.354e-01	0.3342630	0.457843
## 1906	-3.523e-02	7.898e-03	0.204428	2.417e-01	0.2737479	0.349560
## 1907	4.166e-03	-4.243e-02	0.173965	2.213e-01	0.2723556	0.285043
## 1908	-3.112e-02	-1.826e-02	0.224802	2.713e-01	0.2868800	0.439226
## 1909	2.042e-01	-7.922e-02	0.545905	5.592e-01	0.4108985	0.650315
## 1910	9.259e-03	-3.258e-02	0.274669	3.163e-01	0.3296532	0.434981
## 1911	9.517e-03	-3.689e-02	0.163600	1.890e-01	0.2436048	0.277803
## 1912	-5.334e-02	-1.047e-02	0.264223	2.865e-01	0.2889049	0.380857
## 1913	-2.057e-03	-8.990e-02	0.126543	1.697e-01	0.0377590	0.176449
## 2000	8.743e-02	1.546e-02	0.206463	3.221e-01	0.3227226	0.158445
## 2001	-7.692e-02	4.658e-02	0.269924	3.207e-01	0.5180441	0.373912
## 2002	2.194e-03	7.882e-02	0.128483	2.142e-01	0.1957456	0.435388
## 2003	-1.946e-02	-6.000e-03	0.185915	3.484e-01	0.2798991	0.375093
## 2100	3.451e-02	5.088e-02	0.166936	3.327e-02	0.1331556	0.124411
## 2101	NA	NA	NA	NA	NA	NA
## 2102	-9.222e-02	6.725e-02	0.648570	8.381e-01	0.8375249	0.858323
## 2103	-1.258e-01	6.577e-02	0.544664	7.129e-01	0.6684254	0.733985
## 2104	-4.643e-02	7.614e-02	0.305224	3.481e-01	0.4158352	0.615602
## 2105	-6.442e-02	1.421e-01	0.383628	4.964e-01	0.4507399	0.559134
## 2200	6.699e-02	5.767e-02	0.248225	3.191e-01	0.4802056	0.448350
## 2201	-1.180e-01	2.991e-02	0.277255	5.268e-01	0.5473649	0.678631
## 2202	-1.141e-01	5.983e-02	0.494520	5.996e-01	0.6804103	0.476256
## 2203	2.250e-01	-1.996e-03	0.435401	4.517e-01	0.5728298	0.341628
## 2204	-1.634e-02	1.927e-02	0.336842	3.897e-01	0.3759386	0.471432
## 2205	1.182e-01	-7.721e-02	0.521039	6.060e-01	0.6447380	0.627250
## 2206	NA	NA	NA	NA	NA	NA
## 2207	-5.305e-02	-2.348e-02	0.473952	4.794e-01	0.6973184	0.520513
## 2208	-1.092e-02	-7.587e-02	0.368191	3.927e-01	0.4469608	0.482363
## 2209	-1.462e-02	1.932e-02	0.417736	4.620e-01	0.5147533	0.419498
## 2210	2.151e-02	-1.196e-02	0.406396	4.347e-01	0.5288420	0.568908
## 2211	3.199e-02	-1.429e-02	0.298867	3.479e-01	0.3582958	0.422011
## 2212	-2.006e-02	-1.046e-02	0.433045	5.794e-01	0.5301167	0.641385
## 2213	1.068e-01	-2.855e-04	0.202374	2.764e-01	0.2867425	0.259470
## 2214	9.088e-02	1.744e-01	0.316676	3.200e-01	0.3402112	0.511142
## 2215	9.007e-02	-7.401e-02	0.490626	5.626e-01	0.5947743	0.697585
## 2216	2.252e-01	-1.606e-01	0.406333	4.967e-01	0.6960840	0.894988
## 2300	-1.324e-02	6.785e-05	0.313383	3.612e-01	0.4251766	0.431801
## 2301	-3.863e-02	-2.696e-02	0.044826	2.839e-02	0.0052807	0.086134
## 2302	-7.419e-03	-2.070e-02	-0.013595	5.410e-02	0.1012893	0.184882
## 2303	-3.479e-04	-1.811e-02	0.255466	2.701e-01	0.3329765	0.409799
## 2304	-6.712e-03	-2.055e-02	0.170366	1.729e-01	0.1932071	0.242706
## 2305	-3.371e-02	2.388e-02	0.479324	5.139e-01	0.5246450	0.576877
## 2306	-3.938e-03	-4.662e-02	0.198668	2.050e-01	0.1770689	0.315453
## 2307	5.161e-03	-5.835e-02	0.222455	2.154e-01	0.2363070	0.305969
## 2308	7.199e-03	-1.574e-02	0.200791	2.585e-01	0.2329283	0.242832

##	2309	7.110e-02	-1.458e-02	0.185427	1.962e-01	0.2405963	0.286273
##	2310	-7.118e-03	-7.586e-03	0.208709	2.859e-01	0.2819964	0.326001
##	2311	-7.407e-02	1.893e-02	0.320579	3.583e-01	0.3794872	0.484406
##	2312	-2.333e-02	-5.267e-02	0.313817	3.152e-01	0.4005718	0.424811
##	2400	-5.860e-02	-3.756e-02	0.204897	2.344e-01	0.2854255	0.330972
##	2401	-2.108e-01	9.611e-02	-0.159575	-3.798e-02	-0.1159679	0.674752
##	2402	1.133e-02	1.210e-02	0.258688	3.862e-01	0.3941172	0.551323
##	2403	-1.373e-02	-2.863e-04	0.423695	4.767e-01	0.5052620	0.612749
##	2404	-1.278e-02	-3.568e-02	0.235972	2.680e-01	0.3161051	0.380868
##	2405	1.366e-02	-3.261e-02	0.065108	4.379e-02	0.1472753	0.257982
##	2406	-4.799e-02	-6.394e-02	0.189005	1.734e-01	0.2023857	0.270968
##	2500	-5.379e-02	-3.626e-02	0.279031	3.008e-01	0.3747100	0.434517
##	2501	-3.639e-02	-2.416e-01	0.726935	8.551e-01	0.5134966	0.621469
##	2502	-1.535e-02	4.339e-03	0.160414	1.699e-01	0.1677080	0.271613
##	2503	-2.745e-02	7.113e-03	0.712072	7.209e-01	0.7869079	0.760378
##	2504	2.308e-02	1.454e-03	0.098833	1.146e-01	0.1456004	0.173444
##	2505	1.333e-02	-4.959e-02	0.187499	1.775e-01	0.2189759	0.219297
##	2506	-2.901e-02	-3.365e-02	0.672868	7.244e-01	0.7987704	0.778145
##	2507	1.561e-02	-7.632e-02	0.371314	4.187e-01	0.4278750	0.493031
##	2508	-1.266e-03	-9.362e-05	0.306147	3.122e-01	0.3698709	0.368422
##	2600	-1.701e-01	-5.999e-02	0.175228	2.957e-01	0.5058727	0.587437
##	2601	-3.283e-01	1.149e-01	0.290515	5.246e-01	0.5886409	1.083744
##	2602	-1.228e-01	1.105e-01	0.076814	3.123e-01	-0.2063757	-0.123050
##	2603	8.631e-03	-1.096e-01	0.131957	1.900e-01	0.6202743	0.401478
##	2604	-6.425e-02	-2.694e-02	0.248088	2.619e-01	0.2808260	0.290588
##	2605	-1.355e-02	5.000e-03	0.262056	3.069e-01	0.2689967	0.485876
##	2606	9.306e-02	-3.019e-01	0.413073	4.011e-01	0.2731289	0.191243
##	2607	3.792e-02	2.122e-02	0.180763	6.809e-02	0.5310747	0.602631
##	2608	-7.366e-02	-2.034e-01	0.197015	2.766e-01	0.0456885	0.811765
##	2609	-8.503e-02	-1.706e-01	0.202140	6.382e-01	0.3727757	0.113831
##	2610	-5.586e-02	5.798e-02	0.181248	3.252e-01	0.4461168	0.374876
##	2611	-2.206e-02	4.004e-02	0.157398	1.618e-01	0.3145630	0.378259
##	2612	-2.694e-02	-1.818e-01	0.122292	1.124e-01	0.8000908	0.209840
##	2613	-7.660e-02	-2.243e-02	0.285733	3.526e-01	0.5145648	0.600891
##	2614	6.287e-02	-3.297e-02	0.170412	3.346e-01	0.3248532	0.494624
##	2700	5.391e-02	1.618e-02	0.588207	8.155e-01	0.9548862	1.527239
##	2701	1.446e-02	-9.574e-04	0.307470	4.217e-01	0.4608502	0.573959
##	2702	-9.092e-03	-2.165e-02	0.015758	4.472e-02	0.0200173	0.140854
##	2703	5.298e-02	5.055e-02	0.389435	8.459e-01	0.9421067	1.098240
##	2704	-3.173e-02	-4.004e-03	0.344911	3.893e-01	0.4553717	0.547606
##	2705	1.139e-01	7.167e-02	0.284973	3.137e-01	0.4630400	0.711335
##	2706	8.640e-02	2.993e-02	0.198873	5.342e-01	0.8365524	1.086923
##	2707	-2.854e-03	-5.052e-02	0.176367	4.889e-01	0.5995189	0.638049
##	2708	-4.109e-02	-6.376e-02	0.220382	3.616e-01	0.4440190	0.602453
##	2709	3.412e-01	-4.726e-01	-0.067245	-1.354e-03	0.8598441	0.114162
##	2710	3.678e-02	9.027e-02	0.190899	1.861e-01	0.1412537	0.245829
##	2711	3.816e-02	-1.531e-02	0.119285	2.194e-01	0.3831852	0.638088
##	2712	-1.016e-03	-1.012e-02	0.248658	3.652e-01	0.5694920	0.712499
##	2713	1.974e-02	6.887e-03	0.205726	2.544e-01	0.2643463	0.306721
##	2714	1.736e-01	9.677e-02	0.819251	9.714e-01	1.0025867	1.110898

##	2715	9.534e-02	4.045e-03	0.055407	2.718e-01	0.3180273	0.561385
##	2716	-4.225e-02	-6.225e-02	0.148175	2.031e-01	0.1858349	0.307560
##	2717	4.981e-02	3.382e-02	0.200996	3.463e-01	0.4043513	0.469411
##	2718	-1.878e-02	-2.394e-03	0.271495	3.808e-01	0.3986019	0.493904
##	2719	7.217e-02	3.538e-02	0.211450	3.051e-01	0.3484875	0.391242
##	2720	2.133e-02	-5.322e-02	0.054885	1.766e-01	0.2884504	0.480142
##	2721	5.850e-02	3.729e-02	-0.161065	1.226e-01	0.2168326	0.446508
##	2722	1.623e-02	-3.372e-03	0.179960	2.431e-01	0.2635774	0.362007
##	2723	5.859e-03	3.265e-02	0.402072	4.794e-01	0.5250838	0.649230
##	2724	4.119e-02	-3.183e-02	0.210530	3.618e-01	0.6903241	0.902568
##	2725	-1.016e-02	-3.697e-03	0.140435	2.137e-01	0.2449994	0.309230
##	2726	-4.350e-03	-1.846e-02	0.113165	1.509e-01	0.2617293	0.353934
##	2727	2.351e-02	1.712e-02	0.036547	1.388e-01	0.1857399	0.339273
##	2728	4.256e-02	4.083e-02	0.321359	4.203e-01	0.4646186	0.620349
##	2729	6.000e-02	2.829e-02	0.326670	4.665e-01	0.6054014	0.696887
##	2730	-4.305e-03	1.453e-02	0.360428	4.361e-01	0.4710357	0.660694
##	2731	-5.777e-03	-2.661e-02	0.259973	3.285e-01	0.3647591	0.511420
##	2732	6.956e-03	-1.887e-02	0.246227	3.000e-01	0.3411813	0.467486
##	2733	6.336e-02	1.617e-01	0.119465	1.767e-01	0.1985479	0.402695
##	2734	4.034e-03	-1.356e-02	0.209771	1.748e-01	0.3058526	0.433787
##	2735	5.561e-02	4.069e-02	0.241968	2.958e-01	0.4073784	0.541655
##	2736	-7.814e-03	-1.135e-02	0.318348	5.086e-01	0.5296931	0.568106
##	2737	3.892e-02	-3.736e-02	0.032854	6.581e-02	0.0745495	0.261299
##	2738	-4.222e-03	-4.783e-02	0.343865	4.789e-01	0.5950195	0.713497
##	2739	3.419e-02	1.047e-02	0.238117	3.223e-01	0.3893557	0.433151
##	2740	9.926e-02	1.027e-01	0.264410	3.013e-01	0.3670490	0.643172
##	2741	-3.476e-02	-1.001e-01	0.340228	4.606e-01	0.5229320	0.793114
##	2742	-9.216e-02	-1.133e-01	0.345846	4.950e-01	0.5574834	0.588951
##	2743	7.335e-03	-7.661e-04	0.206837	4.199e-01	0.5362701	0.613852
##	2744	3.029e-01	-2.734e-01	0.287830	-1.785e-01	-0.2837513	-0.310715
##	2745	2.008e-02	-4.491e-02	0.250595	4.409e-01	0.4988236	0.605824
##	2746	5.809e-02	7.055e-02	0.158902	2.785e-01	0.3500141	0.602799
##	2747	-7.209e-03	9.702e-03	0.043096	2.075e-01	0.2252531	0.420677
##	2748	7.769e-02	2.280e-01	0.342653	3.887e-01	0.4771719	0.627222
##	2800	-1.725e-02	-1.404e-02	0.338330	4.193e-01	0.4431475	0.536477
##	2801	4.119e-02	1.522e-01	-0.036205	4.846e-02	0.0544628	0.303690
##	2802	6.401e-03	-2.259e-02	0.174917	2.531e-01	0.2772753	0.289151
##	2803	3.536e-02	4.737e-02	0.149326	1.728e-01	0.3027686	0.277232
##	2804	1.066e-02	2.223e-02	0.118837	1.546e-01	0.1687568	0.255708
##	2805	-1.639e-02	1.136e-02	0.187877	2.387e-01	0.2922277	0.347682
##	2806	-1.375e-02	-3.237e-02	0.238128	2.655e-01	0.2884221	0.350524
##	2807	7.751e-02	-2.915e-02	0.049711	2.026e-01	0.3970429	0.274373
##	2808	1.176e-02	1.294e-02	0.215504	2.636e-01	0.3232694	0.400710
##	2809	1.136e-02	-4.653e-02	0.132493	1.934e-01	0.2522919	0.312017
##	2900	1.016e-01	1.604e-02	0.321928	4.831e-01	0.5045446	0.584754
##	2901	2.447e-01	-1.182e-01	-0.056023	3.449e-01	0.1339615	0.299135
##	2902	-1.839e-02	-1.379e-01	0.347364	6.167e-01	0.8096760	0.959831
##	2903	-9.666e-02	-2.269e-01	-0.075716	3.775e-01	0.6245500	0.627159
##	2904	1.933e-02	6.495e-02	-0.029601	3.155e-02	0.1569986	0.170410
##	2905	1.371e-01	-6.052e-02	0.278901	3.247e-01	0.4146539	0.581729

##	2906	1.358e-01	3.194e-01	0.211124	-2.109e-04	0.0002337	0.305946
##	2907	-7.385e-04	-1.209e-01	0.239557	7.214e-01	0.9087513	1.115161
##	2908	-3.722e-01	7.078e-02	0.520299	2.977e-01	0.0764382	1.182859
##	2909	1.973e-02	-8.845e-02	0.320688	4.724e-01	0.5109376	0.597619
##	2910	1.087e-03	-2.147e-02	0.056433	2.014e-01	0.1890860	0.098406
##	2911	5.298e-02	1.225e-01	0.208266	4.317e-01	0.6124688	0.789062
##	2912	-4.944e-02	1.808e-01	0.056291	1.386e-01	0.1698016	0.237236
##	2913	1.317e-01	-1.939e-02	0.378058	5.645e-01	0.5031964	0.671952
##	2914	1.516e-01	-1.465e-01	0.067567	3.321e-01	0.5296935	0.643517
##	2915	NA	NA	NA	NA	NA	NA
##	2916	-2.570e-02	-7.064e-02	0.467121	5.230e-01	0.5703748	0.695890
##	2917	3.529e-03	6.189e-02	0.145621	5.271e-01	0.6817248	0.833955
##	2918	NA	NA	NA	NA	NA	NA
##	2919	-2.652e-02	3.274e-02	0.240570	3.073e-01	0.4570880	0.299143
##	2920	NA	NA	NA	NA	NA	NA
##	2921	1.419e-02	7.258e-02	0.326413	4.000e-01	0.3744098	0.506019
##	2922	-6.919e-02	9.535e-02	0.281185	2.416e-01	0.3675166	0.533805
##	2923	-2.829e-02	-3.776e-02	0.254758	1.502e-01	0.0047442	0.291756
##	3000	-2.998e-02	-9.606e-02	0.491572	6.180e-01	0.6324140	0.633828
##	3001	NA	NA	NA	NA	NA	NA
##	3002	-4.994e-03	-8.201e-03	0.154466	1.845e-01	0.2157774	0.318585
##	3003	9.733e-03	-1.842e-02	0.488284	7.939e-01	0.8537699	0.852906
##	3004	9.718e-05	-2.313e-03	0.383649	4.989e-01	0.5343010	0.580673
##	3005	2.276e-03	-1.587e-02	0.078824	1.165e-01	0.1357445	0.193868
##	3100	-1.753e-02	1.892e-02	0.221393	2.308e-01	0.3090480	0.345869
##	3101	-4.761e-02	-2.651e-02	0.208870	3.312e-01	0.2874462	0.409098
##	3102	5.524e-02	1.182e-02	0.430546	4.252e-01	0.4313195	0.489296
##	3103	-2.472e-02	-5.356e-02	0.317052	3.954e-01	0.3595503	0.403063
##	3104	-1.101e-02	1.199e-02	0.131403	1.274e-01	0.1533273	0.200241
##	3105	1.498e-02	-7.193e-02	0.504968	5.385e-01	0.5605689	0.622389
##	3106	-5.882e-05	-2.992e-02	0.265487	3.836e-01	0.3696804	0.227606
##	3107	-3.601e-02	2.915e-02	0.264644	3.158e-01	0.3959422	0.509258
##	3108	1.008e-01	6.803e-02	0.163015	2.015e-01	0.2799470	0.461804
##	3109	-2.958e-02	-5.355e-03	0.162105	3.150e-01	0.3983221	0.283178
##	3110	4.291e-02	-3.668e-02	0.307072	2.848e-01	0.3981799	0.309633
##	3200	1.558e-02	-1.654e-02	0.270787	3.329e-01	0.3871014	0.420089
##	3201	-3.134e-03	4.465e-02	0.146986	1.035e-01	-0.0903798	0.078850
##	3202	-3.417e-02	-4.113e-02	0.214801	2.427e-01	0.2320396	0.328339
##	3203	-2.468e-02	-2.969e-02	0.372188	5.227e-01	0.5775482	0.650483
##	3204	-2.852e-02	-1.685e-02	0.214346	2.946e-01	0.3444126	0.440195
##	3205	1.446e-02	-1.130e-02	0.177521	2.199e-01	0.2571684	0.312080
##	3206	9.589e-02	4.465e-02	0.332797	4.358e-01	0.4416024	0.468617
##	3207	-5.640e-02	-1.203e-02	0.128267	1.512e-01	0.1931508	0.252239
##	3300	2.971e-02	7.741e-03	0.178729	2.294e-01	0.2552940	0.222616
##	3301	1.786e-02	1.491e-02	0.119326	1.794e-01	0.2141648	0.239400
##	3302	-7.880e-03	-1.066e-01	0.347020	4.781e-01	0.5699942	0.484095
##	3303	6.010e-02	2.485e-02	0.160022	2.633e-01	0.2222478	0.318349
##	3304	4.030e-02	3.682e-02	0.152761	2.010e-01	0.2525286	0.290795
##	3305	5.275e-02	3.836e-02	0.188848	2.652e-01	0.2101993	0.329458
##	3306	2.731e-02	2.524e-02	0.120908	1.844e-01	0.2534336	0.210035



##	3307	7.029e-02	-3.845e-02	0.054117	1.090e-01	0.1195522	0.132692
##	3308	-1.184e-02	6.567e-03	0.171080	2.494e-01	0.3510278	0.271770
##	3309	-1.209e-02	4.667e-03	0.402522	4.808e-01	0.4628103	0.573781
##	3310	3.934e-02	1.216e-02	0.400340	4.753e-01	0.5543310	0.601074
##	3311	5.078e-02	-9.688e-02	0.082585	1.550e-01	-0.0653477	0.080601
##	3312	3.918e-04	3.813e-02	0.167669	1.970e-01	0.2007325	0.225837
##	3313	8.350e-02	4.139e-02	0.337303	3.563e-01	0.3628697	0.269740
##	3314	4.597e-03	6.053e-02	0.244123	2.981e-01	0.1712006	0.134056
##	3315	1.019e-01	-1.275e-02	0.284974	2.605e-01	0.2307610	0.552419
##	3316	2.562e-02	-2.434e-02	0.381264	4.279e-01	0.4834051	0.466404
##	3317	-1.366e-03	8.331e-02	0.160490	1.612e-01	0.4826143	0.423829
##	3318	-5.737e-02	2.764e-02	0.329387	2.854e-01	0.3467228	0.265416
##	3319	6.599e-03	-1.637e-02	0.126255	1.076e-01	0.3079257	0.291877
##	3320	5.499e-02	7.655e-02	0.198318	2.709e-01	0.1774138	0.509380
##	3321	2.234e-02	1.833e-02	0.118814	1.467e-01	0.1527433	0.100866
##	3322	-1.970e-02	-2.713e-02	0.111981	2.419e-01	0.1127023	0.295776
##	3400	1.172e-01	8.963e-03	0.313302	5.098e-01	0.6371817	0.656763
##	3401	-7.103e-02	1.655e-01	0.138857	1.502e-01	0.0785765	0.373240
##	3402	-1.495e-01	-1.783e-02	-0.012221	2.839e-01	0.2778659	-0.419527
##	3403	5.786e-02	4.414e-02	0.050590	-3.852e-02	0.0399830	0.040808
##	3404	1.177e-01	6.976e-02	0.434123	4.810e-01	0.5384690	0.689149
##	3500	8.465e-02	-2.869e-02	0.403614	4.717e-01	0.5233074	0.655347
##	3501	NA	NA	NA	NA	NA	NA
##	3502	NA	NA	NA	NA	NA	NA
##	3503	NA	NA	NA	NA	NA	NA
##	3504	1.233e-01	9.125e-02	0.131539	2.016e-01	0.2118585	0.377015
##	3505	-1.160e-01	7.694e-02	0.314467	4.958e-01	0.4817317	0.589340
##	3506	1.534e-01	-8.980e-02	0.157840	-1.873e-01	-0.0378990	0.001846
##	3600	-5.231e-02	9.580e-02	0.163407	1.237e-01	0.1987595	0.248305
##	3601	-3.641e-02	1.013e-02	0.360371	3.959e-01	0.3823470	0.384164
##	3602	1.103e-01	-1.944e-01	0.282075	3.227e-01	0.2179815	0.255579
##	3603	7.392e-03	-4.326e-02	0.241725	5.408e-01	0.5026066	0.450132
##	3604	NA	NA	NA	NA	NA	NA
##	3605	-8.038e-02	2.064e-01	0.275285	4.385e-01	0.5485781	0.377887
##	3606	NA	NA	NA	NA	NA	NA
##	3607	-1.451e-01	-6.032e-02	0.304531	2.926e-01	0.4114618	0.268825
##	3608	NA	NA	NA	NA	NA	NA
##	3609	5.541e-02	-4.557e-03	0.002375	3.053e-02	0.1082003	0.178098
##	3610	-4.272e-02	-8.279e-02	0.259079	3.426e-01	0.3072324	0.342428
##	3611	2.135e-01	3.643e-02	0.308748	4.778e-01	0.5560899	0.521368
##	3612	-9.117e-02	-1.248e-01	0.295912	3.671e-01	0.4095017	0.480574
##	3613	2.053e-01	8.428e-02	0.155601	3.620e-01	0.5885257	0.606002
##	3614	-1.645e-01	-2.083e-01	0.404605	5.542e-01	0.5914977	0.899444
##	3615	NA	NA	NA	NA	NA	NA
##	3616	4.831e-03	2.441e-02	0.211634	2.335e-01	0.2655576	0.163295
##		FFA2	FLA2	FFA3	FLA4		
##	1000	3.563e-01	-2.319e-01	0.3053362	-9.555e-02		
##	1100	4.341e-02	-2.641e-03	0.0431024	4.457e-03		
##	1101	1.207e-02	1.475e-02	0.0153643	1.868e-02		
##	1102	1.005e-01	1.915e-02	0.1021935	3.371e-02		

## 1103	1.220e-01	5.155e-02	0.1342813	8.876e-02
## 1104	6.370e-03	-4.003e-02	0.0010658	-3.893e-02
## 1105	8.623e-03	5.835e-04	0.0087151	2.468e-03
## 1106	8.811e-02	-1.704e-01	0.0446978	-1.425e-01
## 1107	-1.643e-02	-4.878e-02	-0.0242334	-5.226e-02
## 1108	6.395e-02	1.260e-02	0.0631491	5.639e-03
## 1109	2.116e-02	-3.046e-02	0.0175443	-2.649e-02
## 1110	6.472e-02	-7.946e-02	0.0547435	-6.739e-02
## 1111	-3.004e-02	-3.802e-02	-0.0340472	-4.330e-02
## 1200	4.591e-02	-2.968e-02	0.0193300	1.151e-02
## 1201	3.092e-02	-1.835e-02	0.0210322	-1.140e-03
## 1202	8.727e-03	-2.946e-02	-0.0182393	-2.142e-02
## 1203	7.659e-02	7.091e-03	0.0816453	6.203e-02
## 1204	6.610e-02	-2.178e-01	-0.0848311	-1.688e-01
## 1205	1.726e-01	-1.480e-01	0.0273670	2.074e-02
## 1206	1.872e-01	-1.434e-01	0.0790421	6.523e-04
## 1207	2.635e-01	2.521e-02	0.2813503	2.178e-01
## 1208	5.069e-02	-2.567e-02	0.0259083	2.310e-02
## 1209	5.276e-02	-1.021e-01	-0.0345989	-5.702e-02
## 1210	1.926e-01	-3.164e-02	0.1681907	1.222e-01
## 1211	3.110e-02	-3.279e-02	0.0018141	-4.437e-03
## 1212	1.954e-01	-7.295e-02	0.1269694	1.118e-01
## 1213	8.751e-02	1.655e-02	0.1025029	9.583e-02
## 1300	4.374e-02	-1.601e-02	0.0418373	-8.817e-03
## 1301	9.708e-01	5.985e-01	0.7735751	4.176e-01
## 1302	4.610e-02	7.793e-02	0.0535939	8.332e-02
## 1303	1.265e-02	-5.795e-03	0.0124656	-5.127e-03
## 1304	-2.093e-02	-6.441e-04	-0.0209582	-2.097e-03
## 1305	1.719e-02	-6.138e-02	0.0106873	-5.855e-02
## 1306	-8.424e-03	-4.702e-02	-0.0123056	-4.796e-02
## 1307	3.849e-03	1.598e-03	0.0039617	2.010e-03
## 1308	-8.569e-02	-2.028e-01	-0.1173698	-2.212e-01
## 1309	1.272e-02	3.291e-03	0.0130473	4.831e-03
## 1310	2.147e-02	-1.663e-02	0.0195658	-1.303e-02
## 1311	6.273e-03	-2.218e-02	0.0039128	-2.121e-02
## 1312	7.249e-03	-8.970e-03	0.0068084	-8.410e-03
## 1313	1.157e-02	-2.329e-02	0.0106200	-2.262e-02
## 1314	-1.258e-02	-1.844e-02	-0.0139932	-1.987e-02
## 1315	-4.261e-02	1.532e-02	-0.0420718	1.303e-02
## 1400	-3.621e-02	3.151e-03	-0.0344736	-1.677e-02
## 1401	2.417e-02	3.083e-02	0.0413269	4.432e-02
## 1402	3.503e-02	5.531e-02	0.0639745	7.503e-02
## 1403	-1.938e-02	-2.502e-03	-0.0207285	-1.295e-02
## 1404	7.096e-02	-1.388e-02	0.0634326	2.467e-02
## 1405	7.646e-03	1.106e-02	0.0131542	1.495e-02
## 1406	7.674e-02	2.469e-02	0.0883834	6.123e-02
## 1407	5.576e-02	2.677e-02	0.0715147	5.996e-02
## 1408	-6.654e-03	-1.071e-02	-0.0120777	-1.408e-02
## 1409	3.428e-02	-9.161e-03	0.0293897	8.966e-03
## 1410	9.121e-02	-7.199e-02	0.0482540	-1.580e-02

```
## 1500 -8.873e-03 -5.525e-02 -0.0169056 -5.704e-02
## 1501 -3.618e-01 -6.319e-01 -0.8895743 -8.788e-01
## 1502 8.211e-03 -1.626e-03 0.0079837 -1.712e-05
## 1503 -3.000e-02 -9.089e-03 -0.0300384 -9.283e-03
## 1504 1.056e-01 -3.904e-01 0.0055502 -3.408e-01
## 1505 -3.945e-02 -8.375e-03 -0.0395254 -8.995e-03
## 1506 1.077e-01 1.296e-01 0.1294052 1.530e-01
## 1507 -1.005e-02 5.323e-02 0.0041256 5.000e-02
## 1508 8.172e-02 -1.064e-01 0.0721666 -9.583e-02
## 1600 1.194e-02 1.452e-02 0.0124654 1.518e-02
## 1601 1.157e-01 -1.247e-02 0.1124616 2.131e-02
## 1602 6.448e-02 1.863e-02 0.0656272 2.539e-02
## 1603 -2.106e-02 -2.817e-02 -0.0227135 -3.115e-02
## 1604 1.905e-02 -2.968e-02 0.0173679 -2.812e-02
## 1605 -4.112e-04 -2.924e-02 -0.0005298 -2.925e-02
## 1606 1.851e-02 -4.156e-02 0.0164359 -3.997e-02
## 1607 6.628e-02 -1.345e-03 0.0662063 4.842e-03
## 1700 1.217e-01 6.667e-02 0.1469781 1.157e-01
## 1701 -1.897e-01 3.480e-01 0.0009942 2.536e-01
## 1702 3.671e-02 7.009e-02 0.0638708 8.631e-02
## 1703 3.023e-03 -6.017e-02 -0.0118626 -5.927e-02
## 1704 -1.283e-01 -6.307e-02 -0.1540728 -1.214e-01
## 1705 2.543e-02 -9.672e-02 -0.0164816 -8.432e-02
## 1706 -2.123e-02 -3.314e-02 -0.0340022 -4.243e-02
## 1707 -2.723e-02 5.526e-02 -0.0121477 4.682e-02
## 1708 4.033e-02 -2.006e-02 0.0336639 -3.673e-03
## 1709 4.696e-02 -8.585e-03 0.0435856 1.168e-02
## 1710 1.393e-02 -2.405e-02 0.0021680 -1.657e-02
## 1711 -3.144e-02 1.413e-02 -0.0279711 4.444e-03
## 1712 6.875e-02 1.004e-02 0.0720695 3.564e-02
## 1800 4.042e-02 7.073e-02 0.0756057 9.190e-02
## 1801 NA NA NA NA
## 1802 1.100e-01 2.561e-03 0.1111978 5.768e-02
## 1803 5.741e-02 3.617e-02 0.0726805 6.129e-02
## 1804 -3.097e-02 7.547e-02 -0.0025183 6.173e-02
## 1900 -3.850e-02 -9.203e-02 -0.0847309 -1.127e-01
## 1901 -1.113e-02 -3.116e-03 -0.0118218 -6.266e-03
## 1902 6.489e-02 -2.157e-02 0.0589740 2.402e-03
## 1903 -2.299e-02 -7.719e-03 -0.0266089 -1.988e-02
## 1904 2.169e-02 1.306e-02 0.0278559 2.413e-02
## 1905 5.635e-02 -2.440e-01 0.0209995 -2.296e-01
## 1906 -7.168e-03 1.308e-02 -0.0055679 1.168e-02
## 1907 3.333e-02 -3.628e-02 0.0272468 -2.861e-02
## 1908 -2.719e-03 -1.928e-02 -0.0055750 -1.988e-02
## 1909 3.075e-01 -6.584e-02 0.2838917 6.757e-02
## 1910 5.372e-02 -4.604e-02 0.0442337 -3.079e-02
## 1911 3.396e-02 -4.128e-02 0.0220114 -2.939e-02
## 1912 -6.668e-03 -6.643e-05 -0.0066809 -1.516e-03
## 1913 9.701e-03 -1.025e-01 -0.0173925 -9.954e-02
## 2000 8.157e-02 3.942e-02 0.1074263 9.011e-02
```

##	2001	-2.957e-02	3.499e-02	-0.0078099	1.551e-02
##	2002	1.381e-02	8.366e-02	0.0603351	9.144e-02
##	2003	-3.664e-03	-7.396e-03	-0.0073515	-9.345e-03
##	2100	4.772e-02	4.668e-02	0.0581711	6.149e-02
##	2101	6.002e-01	-2.049e-01	0.4724617	1.695e-01
##	2102	4.266e-02	-4.875e-02	0.0315032	-3.217e-02
##	2103	-1.434e-02	-1.071e-02	-0.0172505	-1.650e-02
##	2104	7.532e-02	7.521e-02	0.0801451	8.385e-02
##	2105	7.591e-03	1.001e-01	0.0490520	1.040e-01
##	2200	1.131e-01	3.933e-02	0.1270287	8.979e-02
##	2201	-6.734e-02	2.695e-02	-0.0564869	-2.264e-03
##	2202	-5.805e-02	4.525e-02	-0.0448220	2.473e-02
##	2203	2.388e-01	-1.048e-02	0.2354746	7.651e-02
##	2204	1.343e-02	1.505e-02	0.0148662	1.696e-02
##	2205	1.364e-01	-4.844e-02	0.1143384	1.308e-02
##	2206	5.358e-02	-2.298e-01	0.0204755	-2.169e-01
##	2207	-5.009e-02	-3.183e-02	-0.0617265	-5.055e-02
##	2208	3.014e-02	-8.290e-02	0.0134250	-7.400e-02
##	2209	-1.687e-02	8.604e-03	-0.0136628	2.277e-03
##	2210	6.666e-02	5.152e-03	0.0677381	2.309e-02
##	2211	9.608e-02	-3.458e-03	0.0953967	2.613e-02
##	2212	3.804e-02	-3.738e-02	0.0273828	-2.323e-02
##	2213	1.586e-01	-9.512e-03	0.1542348	7.412e-02
##	2214	7.677e-02	1.175e-01	0.1434593	1.631e-01
##	2215	1.271e-01	-4.494e-02	0.1083652	9.121e-03
##	2216	3.018e-01	-1.998e-01	0.1838677	-9.715e-03
##	2300	2.410e-02	-1.993e-02	0.0188873	-1.237e-02
##	2301	-3.840e-02	-2.670e-02	-0.0550757	-5.041e-02
##	2302	7.424e-03	-1.366e-02	0.0063324	-1.280e-02
##	2303	2.752e-02	-3.342e-02	0.0234072	-2.877e-02
##	2304	5.526e-03	-2.024e-02	0.0045069	-1.985e-02
##	2305	2.726e-02	5.334e-02	0.0361373	5.903e-02
##	2306	1.865e-02	-6.153e-02	0.0005046	-5.517e-02
##	2307	1.567e-02	-7.019e-02	0.0108866	-6.866e-02
##	2308	2.561e-02	-3.042e-02	0.0091095	-1.575e-02
##	2309	8.933e-02	-3.995e-02	0.0801481	-1.216e-02
##	2310	7.268e-03	-1.183e-02	0.0063558	-1.111e-02
##	2311	-3.885e-02	2.951e-02	-0.0356313	2.356e-02
##	2312	1.097e-02	-7.581e-02	-0.0023542	-7.318e-02
##	2400	-1.927e-02	-2.853e-02	-0.0226274	-3.184e-02
##	2401	-2.181e-01	1.433e-01	-0.1838766	1.425e-01
##	2402	5.337e-02	1.837e-03	0.0534959	8.127e-03
##	2403	7.691e-03	-8.359e-03	0.0073992	-7.947e-03
##	2404	-1.735e-03	-4.287e-02	-0.0027812	-4.294e-02
##	2405	1.731e-02	-5.412e-02	0.0134938	-5.244e-02
##	2406	-4.495e-02	-7.143e-02	-0.0501592	-7.664e-02
##	2500	-5.645e-03	-3.085e-02	-0.0093036	-3.200e-02
##	2501	-1.212e-02	-3.677e-01	-0.2279287	-3.750e-01
##	2502	5.952e-03	7.386e-03	0.0066426	8.296e-03
##	2503	1.174e-03	2.722e-02	0.0016992	2.726e-02

## 2504	3.536e-02	8.216e-03	0.0358292	1.180e-02
## 2505	1.980e-02	-4.890e-02	0.0176570	-4.749e-02
## 2506	2.282e-02	7.358e-04	0.0228281	1.662e-03
## 2507	4.780e-02	-6.341e-02	0.0420528	-5.593e-02
## 2508	1.383e-02	7.034e-03	0.0140524	7.826e-03
## 2600	-1.302e-01	-7.629e-02	-0.1739749	-1.614e-01
## 2601	-2.669e-01	7.999e-02	-0.2199858	-9.413e-02
## 2602	-1.144e-01	1.093e-01	-0.0371992	2.755e-02
## 2603	3.662e-02	-1.232e-01	-0.0352172	-9.963e-02
## 2604	-2.813e-02	-2.509e-02	-0.0361558	-3.590e-02
## 2605	6.043e-02	9.359e-03	0.0619642	2.300e-02
## 2606	7.420e-02	-2.397e-01	-0.0445590	-2.086e-01
## 2607	4.716e-02	1.749e-02	0.0555046	4.089e-02
## 2608	-1.073e-02	-2.491e-01	-0.1806711	-2.577e-01
## 2609	6.213e-02	-2.074e-01	-0.0311276	-1.661e-01
## 2610	-1.088e-02	2.069e-02	-0.0021449	1.448e-02
## 2611	8.900e-03	4.574e-02	0.0186872	4.821e-02
## 2612	-1.988e-02	-1.881e-01	-0.1055559	-1.928e-01
## 2613	-1.405e-02	-1.270e-02	-0.0178130	-1.794e-02
## 2614	8.925e-02	-1.910e-02	0.0819225	1.863e-02
## 2700	1.758e-01	-1.675e-01	0.0954397	-7.403e-02
## 2701	6.385e-02	-5.963e-02	0.0445858	-3.723e-02
## 2702	2.482e-03	-3.065e-02	-0.0060689	-2.982e-02
## 2703	1.348e-01	2.684e-02	0.1449519	8.521e-02
## 2704	-1.616e-02	-2.164e-03	-0.0163551	-3.781e-03
## 2705	1.382e-01	4.730e-02	0.1435031	7.509e-02
## 2706	1.640e-01	-3.677e-02	0.1534788	2.325e-02
## 2707	8.011e-02	-1.073e-01	0.0246116	-6.333e-02
## 2708	-1.103e-02	-1.105e-01	-0.0371350	-1.135e-01
## 2709	3.233e-01	-4.551e-01	-0.0175884	-2.166e-01
## 2710	5.636e-02	9.632e-02	0.0723584	1.079e-01
## 2711	3.664e-02	-1.348e-02	0.0334401	-3.025e-03
## 2712	4.258e-02	-7.310e-02	0.0330463	-6.491e-02
## 2713	5.712e-02	1.837e-02	0.0610378	3.209e-02
## 2714	2.843e-01	8.815e-02	0.3014328	1.499e-01
## 2715	1.400e-01	-1.978e-03	0.1398595	2.072e-02
## 2716	-1.610e-02	-7.487e-02	-0.0273657	-7.781e-02
## 2717	7.230e-02	-1.321e-02	0.0688188	7.241e-03
## 2718	-2.099e-03	-2.484e-02	-0.0089856	-2.546e-02
## 2719	1.285e-01	2.694e-02	0.1396775	8.210e-02
## 2720	5.159e-02	-5.966e-02	0.0462994	-5.329e-02
## 2721	8.650e-02	2.435e-02	0.0877639	3.480e-02
## 2722	5.303e-02	-9.253e-03	0.0512869	3.877e-03
## 2723	4.949e-02	3.365e-02	0.0514321	3.734e-02
## 2724	1.124e-01	-1.078e-01	0.0977803	-8.423e-02
## 2725	1.327e-02	-1.279e-02	0.0115964	-1.053e-02
## 2726	2.571e-02	-1.562e-02	0.0242539	-1.226e-02
## 2727	4.116e-02	3.437e-03	0.0417190	1.340e-02
## 2728	7.529e-02	2.004e-02	0.0776433	3.177e-02
## 2729	8.353e-02	-5.928e-02	0.0655140	-3.188e-02

##	2730	2.400e-02	-2.856e-02	0.0196438	-2.385e-02
##	2731	1.363e-02	-5.063e-02	0.0078489	-4.850e-02
##	2732	-1.162e-03	-5.015e-02	-0.0153831	-5.056e-02
##	2733	6.232e-02	1.579e-01	0.0974482	1.765e-01
##	2734	3.835e-02	-4.429e-02	0.0274286	-3.386e-02
##	2735	7.434e-02	-9.259e-03	0.0717619	1.320e-02
##	2736	5.273e-02	-3.440e-03	0.0523784	4.371e-03
##	2737	3.638e-02	-5.784e-02	0.0308594	-5.244e-02
##	2738	5.344e-02	-8.589e-02	0.0311527	-7.113e-02
##	2739	7.609e-02	-1.933e-02	0.0698469	6.434e-03
##	2740	1.290e-01	1.027e-01	0.1417096	1.289e-01
##	2741	-1.794e-02	-1.927e-01	-0.0547005	-1.977e-01
##	2742	-7.790e-02	-1.814e-01	-0.1499612	-2.115e-01
##	2743	3.398e-02	-1.044e-01	-0.0029741	-9.191e-02
##	2744	3.479e-01	-3.359e-01	0.0911159	-3.225e-02
##	2745	3.898e-02	-6.946e-02	0.0316529	-6.430e-02
##	2746	5.994e-02	2.838e-02	0.0637430	4.163e-02
##	2747	1.295e-05	-3.875e-03	-0.0003749	-3.868e-03
##	2748	9.328e-02	2.043e-01	0.1195147	2.233e-01
##	2800	9.922e-03	-2.230e-02	0.0084213	-2.133e-02
##	2801	2.738e-02	1.502e-01	0.0739866	1.604e-01
##	2802	2.144e-02	-3.262e-02	0.0185796	-3.040e-02
##	2803	4.347e-02	4.591e-02	0.0464904	4.889e-02
##	2804	2.219e-02	1.794e-02	0.0238506	2.077e-02
##	2805	9.876e-03	-1.162e-03	0.0096873	8.770e-04
##	2806	9.424e-03	-4.252e-02	0.0007712	-4.037e-02
##	2807	6.553e-02	3.690e-03	0.0657155	6.974e-03
##	2808	2.729e-02	1.806e-03	0.0274430	4.983e-03
##	2809	3.358e-02	-5.597e-02	0.0262341	-5.002e-02
##	2900	1.542e-01	-2.840e-02	0.1372322	6.004e-02
##	2901	3.386e-01	-1.628e-01	0.2622252	-2.162e-02
##	2902	6.596e-02	-4.342e-01	-0.1911116	-3.964e-01
##	2903	-1.075e-01	-7.049e-01	-0.4183722	-7.636e-01
##	2904	3.711e-02	6.921e-02	0.0536236	7.816e-02
##	2905	1.360e-01	-4.433e-02	0.1048288	5.133e-02
##	2906	1.549e-01	2.942e-01	0.3327241	3.774e-01
##	2907	-1.117e-01	-3.162e-01	-0.3226810	-3.927e-01
##	2908	-1.465e-01	-1.424e-01	-0.2457864	-2.374e-01
##	2909	8.297e-02	-1.777e-01	-0.0130857	-1.331e-01
##	2910	1.064e-02	-1.622e-02	-0.0001513	-9.066e-03
##	2911	1.692e-01	5.010e-02	0.1929669	1.329e-01
##	2912	-3.340e-02	1.923e-01	0.0309801	1.821e-01
##	2913	3.602e-01	-2.012e-01	0.2134457	-4.422e-03
##	2914	NA	NA	NA	NA
##	2915	NA	NA	NA	NA
##	2916	-6.266e-03	-1.474e-01	-0.0450712	-1.492e-01
##	2917	1.008e-01	3.258e-02	0.1147539	7.492e-02
##	2918	NA	NA	NA	NA
##	2919	-2.346e-03	2.404e-02	0.0070611	2.318e-02
##	2920	NA	NA	NA	NA

##	2921	5.561e-02	9.427e-02	0.1010017	1.206e-01
##	2922	-3.197e-02	1.067e-01	0.0400115	8.560e-02
##	2923	-4.598e-02	-1.733e-02	-0.0560661	-4.170e-02
##	3000	4.536e-02	-7.647e-02	0.0401425	-7.159e-02
##	3001	NA	NA	NA	NA
##	3002	3.077e-03	5.594e-04	0.0030753	5.455e-04
##	3003	2.126e-02	-1.135e-01	-0.0029175	-1.075e-01
##	3004	5.066e-02	-1.083e-02	0.0499209	-5.552e-03
##	3005	7.129e-03	-2.457e-02	0.0049669	-2.378e-02
##	3100	2.757e-02	8.831e-03	0.0294135	1.770e-02
##	3101	-1.693e-02	-4.577e-02	-0.0319676	-5.251e-02
##	3102	1.186e-01	2.001e-02	0.1215683	4.727e-02
##	3103	3.326e-02	-5.637e-02	0.0207551	-4.591e-02
##	3104	9.347e-03	1.544e-02	0.0108173	1.705e-02
##	3105	9.266e-02	-7.187e-02	0.0770635	-4.157e-02
##	3106	1.283e-02	-3.224e-02	0.0029436	-2.749e-02
##	3107	1.723e-02	3.312e-02	0.0217699	3.735e-02
##	3108	1.866e-01	1.179e-01	0.2074021	1.709e-01
##	3109	8.614e-03	-3.638e-02	-0.0053371	-3.192e-02
##	3110	6.956e-02	-3.545e-02	0.0684866	-3.117e-02
##	3200	3.732e-02	-1.772e-02	0.0315669	-5.234e-03
##	3201	1.707e-02	5.355e-02	0.0397215	6.094e-02
##	3202	-1.986e-02	-5.135e-02	-0.0356736	-5.771e-02
##	3203	3.685e-02	-5.726e-02	0.0190098	-4.564e-02
##	3204	3.446e-04	-2.315e-02	-0.0075213	-2.303e-02
##	3205	3.562e-02	-8.268e-03	0.0341922	-1.630e-03
##	3206	1.339e-01	4.021e-02	0.1377889	5.443e-02
##	3207	-4.845e-02	-1.803e-02	-0.0573320	-4.206e-02
##	3300	4.728e-02	-6.256e-04	0.0468364	3.331e-02
##	3301	3.637e-02	1.367e-02	0.0454546	3.791e-02
##	3302	6.605e-02	-1.993e-01	-0.0852013	-1.469e-01
##	3303	5.858e-02	2.846e-02	0.0794146	7.126e-02
##	3304	5.359e-02	3.904e-02	0.0765704	7.071e-02
##	3305	5.709e-02	2.981e-02	0.0786728	7.147e-02
##	3306	4.927e-02	1.153e-02	0.0554984	3.818e-02
##	3307	8.273e-02	-3.786e-02	0.0733478	-1.406e-02
##	3308	6.078e-03	1.727e-03	0.0073961	6.413e-03
##	3309	5.240e-02	-2.739e-02	0.0340601	8.734e-03
##	3310	9.013e-02	-3.202e-03	0.0878297	6.208e-02
##	3311	6.467e-02	-1.036e-01	-0.0076894	-5.583e-02
##	3312	1.263e-02	3.694e-02	0.0404138	4.652e-02
##	3313	9.121e-02	4.575e-02	0.1135214	9.021e-02
##	3314	2.908e-02	4.519e-02	0.0666637	6.962e-02
##	3315	1.013e-01	-1.733e-02	0.0879636	6.082e-02
##	3316	3.511e-02	-2.721e-02	0.0107245	4.308e-03
##	3317	2.383e-02	5.688e-02	0.0658151	7.486e-02
##	3318	-1.682e-02	-3.471e-02	-0.0458630	-4.809e-02
##	3319	2.297e-02	-3.219e-02	0.0048382	-1.940e-02
##	3320	3.505e-02	9.967e-02	0.1213037	1.295e-01
##	3321	2.865e-02	2.560e-02	0.0442161	4.301e-02

```

## 3322 -2.591e-02 -2.454e-02 -0.0419451 -4.144e-02
## 3400 2.301e-01 -4.064e-02 0.2165248 4.859e-02
## 3401 -1.165e-01 1.144e-01 -0.0986735 1.031e-01
## 3402 -1.505e-01 3.869e-02 -0.1249161 -6.070e-02
## 3403 6.384e-02 3.827e-02 0.0683541 4.601e-02
## 3404 2.434e-01 1.333e-01 0.2777160 1.930e-01
## 3500 1.402e-01 -8.947e-02 0.1111496 -3.294e-02
## 3501 NA NA NA NA
## 3502 NA NA NA NA
## 3503 NA NA NA NA
## 3504 1.382e-01 1.009e-01 0.1560284 1.355e-01
## 3505 -5.465e-02 3.995e-02 -0.0481373 2.637e-02
## 3506 1.655e-01 -8.298e-02 0.1575175 -5.213e-02
## 3600 -4.992e-02 7.659e-02 -0.0106130 5.045e-02
## 3601 -1.636e-02 1.379e-02 -0.0084481 4.492e-03
## 3602 1.224e-01 -2.311e-01 0.0177718 -1.800e-01
## 3603 3.747e-02 -3.496e-02 0.0261185 -1.307e-02
## 3604 NA NA NA NA
## 3605 -5.903e-02 1.851e-01 0.0208091 1.586e-01
## 3606 NA NA NA NA
## 3607 -1.220e-01 -7.410e-02 -0.1448983 -1.171e-01
## 3608 NA NA NA NA
## 3609 6.207e-02 -1.504e-02 0.0547558 8.760e-03
## 3610 3.416e-04 -7.527e-02 -0.0099862 -7.523e-02
## 3611 2.709e-01 2.057e-02 0.2765709 8.899e-02
## 3612 -7.547e-02 -1.618e-01 -0.1472231 -1.968e-01
## 3613 2.292e-01 8.922e-02 0.2448107 1.632e-01
## 3614 -1.707e-01 -4.000e-01 -0.2947835 -4.789e-01
## 3615 NA NA NA NA
## 3616 2.195e-02 1.100e-02 0.0258196 1.825e-02

```

```
print(RegP)
```

```

##          FFA1p      FLA1p          2p          3p          4p          5+p
## 1000 8.102e-01 7.347e-02 7.125e-238 0.000e+00 0.000e+00 0.000e+00
## 1100 2.959e-01 6.150e-01 1.871e-41 3.424e-64 4.525e-64 1.012e-95
## 1101 2.676e-01 9.708e-01 4.402e-06 3.698e-05 3.941e-05 6.607e-09
## 1102 2.400e-02 4.189e-01 2.185e-06 3.925e-16 6.899e-21 1.460e-28
## 1103 1.331e-06 3.723e-04 1.175e-27 7.974e-36 2.998e-35 1.538e-47
## 1104 5.701e-01 5.157e-02 3.142e-08 1.315e-12 2.600e-15 6.850e-29
## 1105 5.982e-02 3.008e-01 4.133e-37 1.861e-49 1.751e-49 6.274e-85
## 1106 3.326e-02 1.915e-04 6.292e-33 3.568e-59 4.051e-63 1.623e-84
## 1107 4.981e-01 3.170e-01 2.282e-05 1.610e-02 2.287e-03 1.648e-05
## 1108 4.309e-01 8.411e-01 3.101e-02 9.095e-03 2.717e-03 1.746e-04
## 1109 2.358e-01 5.519e-01 9.615e-19 5.657e-26 5.160e-27 1.461e-46
## 1110 1.836e-01 3.812e-02 5.653e-30 9.877e-49 2.739e-66 5.394e-91
## 1111 1.682e-01 2.237e-01 2.279e-01 4.626e-01 1.012e-01 8.907e-03
## 1200 2.677e-01 3.241e-01 3.293e-06 3.236e-13 1.075e-03 1.742e-10
## 1201 2.466e-01 9.625e-01 1.311e-32 9.465e-30 3.582e-18 4.958e-30
## 1202 4.658e-01 7.838e-01 1.869e-26 6.147e-09 3.260e-20 7.092e-02

```



##	1203	1.498e-01	2.967e-01	7.871e-71	1.078e-75	3.062e-49	3.314e-38
##	1204	9.432e-01	6.519e-03	1.054e-12	7.470e-16	9.134e-14	9.416e-12
##	1205	7.336e-01	7.659e-01	5.497e-01	1.849e-01	8.646e-01	9.864e-13
##	1206	5.358e-01	5.825e-01	5.606e-02	7.287e-04	3.610e-02	3.682e-01
##	1207	2.427e-07	9.853e-02	1.393e-33	1.492e-49	2.360e-23	1.187e-44
##	1208	3.822e-01	5.319e-01	2.403e-06	1.325e-03	8.109e-02	9.039e-01
##	1209	8.349e-01	6.879e-01	1.111e-01	2.308e-01	7.692e-03	4.250e-04
##	1210	2.024e-01	4.905e-01	3.306e-10	6.892e-03	6.013e-05	1.451e-01
##	1211	8.810e-01	8.903e-01	7.703e-07	5.480e-04	5.945e-01	7.971e-01
##	1212	3.058e-01	8.578e-01	7.885e-07	1.035e-02	8.217e-02	2.990e-01
##	1213	2.464e-01	8.105e-01	1.319e-02	1.184e-01	4.914e-03	7.495e-01
##	1300	7.497e-01	3.238e-01	4.260e-29	5.197e-32	7.996e-35	2.563e-77
##	1301	NA	NA	NA	NA	NA	NA
##	1302	3.582e-01	2.090e-02	1.304e-01	9.638e-04	8.077e-04	1.400e-07
##	1303	5.467e-01	2.250e-01	3.955e-14	5.114e-16	7.527e-21	1.084e-39
##	1304	3.673e-03	7.993e-01	2.771e-07	5.396e-07	9.525e-08	8.696e-14
##	1305	1.321e-01	4.661e-02	7.025e-26	4.475e-35	2.038e-40	7.710e-64
##	1306	1.452e-03	1.708e-02	1.736e-43	1.743e-51	7.522e-50	2.622e-101
##	1307	5.971e-01	5.670e-01	1.087e-10	5.818e-14	1.045e-15	1.195e-28
##	1308	9.383e-06	8.376e-11	3.652e-13	6.751e-37	1.058e-56	9.618e-126
##	1309	7.975e-01	8.040e-01	2.720e-04	6.792e-04	6.805e-05	3.864e-08
##	1310	9.618e-01	6.390e-01	2.861e-11	4.880e-13	5.357e-21	5.433e-35
##	1311	3.555e-03	1.541e-01	4.497e-36	9.339e-47	2.044e-53	1.859e-87
##	1312	3.081e-01	4.882e-01	4.843e-36	1.804e-41	4.535e-51	3.777e-78
##	1313	5.078e-01	6.762e-02	4.909e-08	5.792e-26	3.005e-25	2.110e-45
##	1314	5.034e-02	2.503e-01	1.401e-02	7.875e-08	6.521e-08	2.687e-17
##	1315	7.965e-03	3.076e-01	8.810e-01	5.897e-01	4.096e-01	2.798e-02
##	1400	1.204e-01	8.189e-01	3.092e-17	1.535e-23	4.834e-11	1.886e-02
##	1401	6.831e-01	5.617e-01	3.812e-11	1.092e-08	9.330e-05	2.697e-03
##	1402	3.068e-01	4.641e-02	1.661e-02	8.696e-03	1.884e-03	5.573e-04
##	1403	2.186e-01	9.262e-01	3.478e-15	1.287e-26	9.335e-07	5.179e-01
##	1404	1.314e-01	8.088e-01	3.564e-04	1.410e-03	9.159e-03	3.365e-01
##	1405	9.498e-01	7.217e-01	1.001e-15	2.723e-19	3.077e-08	4.311e-02
##	1406	2.760e-02	4.799e-01	3.630e-18	6.733e-16	1.264e-08	2.800e-02
##	1407	3.020e-02	2.826e-01	2.132e-29	7.100e-17	2.516e-15	5.649e-07
##	1408	4.339e-01	3.997e-01	4.094e-34	1.287e-31	1.853e-14	6.232e-06
##	1409	2.587e-01	7.335e-01	9.596e-02	1.566e-01	2.062e-01	3.460e-01
##	1410	3.632e-01	3.171e-01	1.148e-07	2.722e-06	1.170e-02	1.784e-04
##	1500	5.484e-02	2.964e-02	5.176e-28	4.885e-29	3.819e-31	4.245e-40
##	1501	2.022e-02	2.828e-01	5.910e-01	2.010e-05	1.118e-05	3.162e-04
##	1502	2.145e-01	6.619e-01	6.553e-17	1.311e-23	5.437e-27	1.172e-44
##	1503	6.775e-03	4.380e-01	1.353e-01	1.310e-01	1.419e-03	1.424e-03
##	1504	6.158e-01	2.232e-02	9.240e-03	2.885e-01	8.213e-01	4.082e-01
##	1505	4.869e-02	8.194e-01	2.212e-01	2.631e-02	1.714e-03	2.998e-04
##	1506	2.996e-01	2.097e-01	2.502e-02	1.582e-04	5.043e-02	7.216e-02
##	1507	5.596e-01	4.918e-01	2.198e-01	4.387e-02	1.678e-02	8.312e-03
##	1508	4.145e-01	2.051e-01	7.087e-13	2.400e-31	5.534e-34	1.719e-40
##	1600	3.033e-01	8.267e-01	1.104e-38	1.092e-44	1.652e-53	1.557e-67
##	1601	1.668e-01	7.852e-01	6.090e-06	1.597e-05	2.157e-03	5.302e-07
##	1602	1.541e-02	4.815e-01	1.266e-10	3.463e-13	7.851e-17	3.606e-22

## 1603	2.647e-01	5.435e-01	6.833e-02	1.709e-01	3.730e-02	3.609e-02
## 1604	2.700e-01	1.243e-01	5.745e-01	4.734e-01	1.237e-01	2.367e-02
## 1605	7.874e-01	3.677e-03	2.013e-01	2.610e-02	7.409e-04	2.069e-06
## 1606	1.997e-01	5.482e-04	4.754e-07	5.286e-05	7.724e-08	1.305e-10
## 1607	1.154e-02	8.996e-01	9.602e-06	5.383e-08	1.148e-09	4.326e-10
## 1700	9.469e-02	3.730e-01	3.529e-08	4.159e-07	5.700e-07	8.425e-13
## 1701	2.658e-01	4.183e-02	2.626e-02	9.623e-02	5.331e-01	3.064e-01
## 1702	5.440e-01	1.287e-02	3.757e-10	5.826e-09	1.499e-10	3.082e-06
## 1703	2.167e-01	1.927e-01	3.143e-13	2.757e-16	3.286e-12	2.481e-25
## 1704	1.001e-01	9.446e-01	1.178e-05	3.123e-11	6.735e-07	1.729e-07
## 1705	7.453e-01	2.939e-01	7.261e-08	1.794e-15	1.214e-13	1.122e-09
## 1706	2.867e-01	7.909e-01	1.647e-21	5.390e-24	3.541e-27	2.390e-27
## 1707	5.389e-01	2.580e-01	4.047e-02	1.041e-02	3.639e-02	1.295e-03
## 1708	7.930e-01	9.848e-01	1.479e-12	1.087e-08	5.461e-12	1.229e-09
## 1709	5.913e-01	6.063e-01	3.021e-11	1.186e-20	1.392e-10	4.092e-10
## 1710	9.970e-01	9.165e-01	2.316e-09	1.579e-18	1.852e-11	4.594e-09
## 1711	7.239e-01	6.309e-01	3.007e-02	3.152e-02	2.659e-02	4.742e-04
## 1712	9.122e-02	4.867e-01	4.750e-23	4.898e-24	6.590e-23	5.582e-20
## 1800	6.274e-01	1.273e-01	8.477e-02	3.673e-03	4.089e-01	2.394e-02
## 1801	NA	NA	NA	NA	NA	NA
## 1802	1.315e-02	9.094e-01	6.677e-02	9.992e-01	1.428e-01	1.957e-02
## 1803	2.357e-01	3.815e-01	4.993e-05	7.915e-08	4.531e-07	7.042e-02
## 1804	1.708e-01	1.636e-01	7.096e-08	8.619e-11	1.425e-06	1.891e-06
## 1900	1.050e-01	3.769e-01	2.347e-10	9.773e-12	1.098e-07	1.916e-22
## 1901	1.586e-01	7.508e-01	2.230e-13	4.354e-16	2.069e-15	3.626e-22
## 1902	8.244e-01	7.959e-01	1.646e-45	7.072e-52	3.865e-49	1.364e-72
## 1903	3.195e-01	4.559e-01	2.634e-01	2.832e-03	5.835e-02	9.955e-06
## 1904	5.620e-01	2.042e-01	3.347e-13	7.885e-14	1.279e-08	5.961e-18
## 1905	3.702e-01	1.046e-02	6.489e-02	1.684e-05	4.067e-04	3.098e-04
## 1906	9.277e-02	7.787e-01	4.139e-12	3.755e-16	4.950e-18	8.872e-26
## 1907	8.770e-01	2.187e-01	8.223e-09	1.900e-12	5.454e-13	3.277e-14
## 1908	1.883e-01	5.533e-01	3.531e-15	2.585e-18	6.055e-17	1.270e-33
## 1909	4.921e-02	4.484e-01	1.780e-10	5.090e-10	3.792e-03	6.324e-12
## 1910	6.578e-01	2.043e-01	1.092e-20	1.397e-25	4.430e-23	1.193e-41
## 1911	6.779e-01	1.463e-01	7.637e-12	6.764e-13	4.256e-14	3.962e-17
## 1912	1.459e-02	7.023e-01	3.028e-25	5.975e-25	2.021e-18	1.407e-25
## 1913	9.686e-01	1.384e-01	4.048e-02	1.304e-02	6.445e-01	4.637e-02
## 2000	1.443e-01	7.887e-01	2.918e-06	5.968e-07	4.149e-03	4.728e-01
## 2001	2.493e-01	4.922e-01	4.699e-06	5.126e-07	2.389e-03	6.870e-03
## 2002	9.002e-01	7.759e-06	2.080e-21	2.501e-28	1.884e-07	7.212e-12
## 2003	4.880e-01	8.419e-01	1.712e-16	2.759e-33	3.752e-05	1.675e-02
## 2100	3.827e-01	3.490e-01	4.718e-03	5.852e-01	3.282e-02	2.631e-02
## 2101	NA	NA	NA	NA	NA	NA
## 2102	1.822e-01	4.728e-01	2.513e-20	1.245e-38	1.965e-26	2.665e-32
## 2103	4.261e-02	4.369e-01	1.067e-13	9.093e-23	2.498e-15	2.646e-23
## 2104	6.882e-01	6.724e-01	4.654e-03	1.857e-03	1.392e-03	1.380e-05
## 2105	1.951e-01	1.483e-02	3.749e-14	3.607e-19	1.263e-12	8.510e-23
## 2200	7.461e-02	1.912e-01	6.092e-12	2.168e-15	1.286e-20	6.019e-18
## 2201	8.273e-02	6.735e-01	6.129e-05	3.398e-13	8.056e-09	3.407e-13
## 2202	1.648e-01	4.828e-01	4.027e-19	3.032e-19	2.070e-16	2.277e-05

##	2203	4.532e-02	9.877e-01	4.436e-06	3.361e-06	4.862e-06	2.438e-02
##	2204	5.078e-01	5.314e-01	3.370e-11	1.955e-16	7.161e-15	1.984e-24
##	2205	1.419e-02	1.464e-01	1.237e-49	1.663e-44	6.775e-31	3.101e-24
##	2206	NA	NA	NA	NA	NA	NA
##	2207	3.897e-01	6.954e-01	2.136e-27	1.183e-23	2.116e-28	1.462e-12
##	2208	7.050e-01	5.478e-02	2.053e-41	1.289e-42	5.774e-40	3.352e-50
##	2209	7.251e-01	6.256e-01	2.243e-25	2.393e-30	8.916e-24	6.480e-14
##	2210	5.956e-01	8.037e-01	6.757e-34	6.165e-34	8.349e-37	4.198e-35
##	2211	3.138e-01	7.306e-01	9.206e-17	7.568e-22	3.749e-18	3.993e-23
##	2212	7.915e-01	9.172e-01	1.866e-09	4.871e-12	9.819e-07	1.809e-08
##	2213	1.880e-02	9.954e-01	3.646e-06	1.450e-07	9.391e-06	4.345e-04
##	2214	3.995e-01	1.427e-01	6.828e-05	1.741e-03	3.616e-02	7.385e-04
##	2215	9.339e-02	2.080e-01	7.079e-27	9.716e-22	2.894e-13	7.753e-16
##	2216	2.803e-03	5.410e-02	1.056e-09	3.057e-08	7.138e-10	8.791e-08
##	2300	4.865e-01	9.974e-01	2.793e-37	8.786e-42	4.104e-45	1.789e-54
##	2301	1.703e-01	3.434e-01	5.789e-02	3.823e-01	9.349e-01	2.283e-01
##	2302	8.815e-01	7.106e-01	8.256e-01	4.392e-01	1.785e-01	7.927e-03
##	2303	9.809e-01	2.840e-01	3.498e-24	3.636e-26	9.917e-37	9.163e-56
##	2304	6.737e-01	2.736e-01	2.715e-06	1.534e-06	1.264e-07	1.140e-11
##	2305	2.946e-01	5.017e-01	1.099e-22	3.708e-25	5.089e-23	2.639e-30
##	2306	9.127e-01	2.412e-01	2.164e-06	3.241e-05	1.539e-03	9.823e-12
##	2307	8.351e-01	3.808e-02	2.546e-04	2.098e-04	1.068e-04	1.004e-07
##	2308	6.995e-01	4.173e-01	1.287e-29	6.741e-29	4.957e-14	2.755e-13
##	2309	5.753e-03	6.381e-01	2.250e-07	7.139e-08	1.519e-09	1.355e-12
##	2310	7.596e-01	7.638e-01	1.058e-05	6.140e-10	3.689e-09	1.368e-12
##	2311	2.447e-02	6.067e-01	1.148e-09	2.116e-11	6.754e-12	4.406e-20
##	2312	3.544e-01	8.346e-02	2.229e-21	2.915e-18	2.236e-23	3.361e-27
##	2400	5.533e-03	1.339e-01	2.347e-07	5.780e-09	1.668e-11	1.394e-18
##	2401	7.290e-02	3.547e-01	4.541e-01	7.017e-01	3.198e-01	5.368e-02
##	2402	6.390e-01	6.933e-01	1.472e-05	5.336e-11	1.122e-11	3.055e-22
##	2403	1.312e-01	9.791e-01	2.392e-21	3.955e-27	9.207e-31	1.580e-45
##	2404	2.263e-01	1.682e-02	2.187e-06	4.932e-08	1.165e-10	6.328e-15
##	2405	4.387e-01	1.155e-01	2.018e-01	3.843e-01	2.702e-03	4.940e-08
##	2406	1.434e-03	6.696e-04	6.120e-03	1.035e-02	2.765e-03	5.220e-05
##	2500	2.419e-02	2.794e-01	4.197e-18	1.349e-21	5.048e-28	1.154e-39
##	2501	7.990e-01	1.861e-01	2.727e-04	1.977e-05	7.265e-03	6.875e-05
##	2502	5.079e-01	8.835e-01	5.251e-03	2.261e-03	3.391e-03	9.065e-07
##	2503	1.138e-01	7.586e-01	6.270e-35	1.741e-38	2.062e-45	1.112e-42
##	2504	1.681e-01	9.484e-01	9.532e-03	2.384e-03	1.507e-04	2.663e-06
##	2505	2.761e-01	1.782e-03	5.101e-08	1.449e-07	1.232e-10	6.327e-11
##	2506	1.302e-01	1.734e-01	8.347e-48	5.178e-66	3.062e-80	4.780e-77
##	2507	5.925e-01	6.747e-02	2.043e-11	1.661e-14	8.350e-14	4.059e-18
##	2508	9.330e-01	9.963e-01	1.369e-10	2.942e-11	4.886e-15	2.682e-15
##	2600	6.303e-04	2.681e-01	3.590e-08	2.118e-09	7.827e-10	4.926e-16
##	2601	1.376e-02	4.039e-01	4.147e-03	3.241e-05	8.929e-02	1.099e-01
##	2602	1.144e-01	1.655e-01	1.015e-01	1.572e-02	4.849e-01	4.346e-01
##	2603	9.283e-01	2.700e-01	2.198e-02	9.767e-02	1.076e-02	1.162e-03
##	2604	2.979e-02	3.986e-01	9.613e-28	6.754e-20	6.759e-12	1.030e-15
##	2605	7.837e-01	9.381e-01	1.694e-08	3.954e-09	1.018e-03	1.054e-09
##	2606	4.860e-01	8.323e-03	1.471e-07	1.471e-03	1.114e-01	3.069e-01

##	2607	7.202e-01	8.383e-01	3.200e-03	5.318e-01	2.063e-04	1.684e-03
##	2608	5.880e-01	1.498e-01	6.905e-04	1.209e-01	7.862e-01	7.143e-23
##	2609	4.379e-01	2.036e-01	7.951e-03	1.106e-04	7.990e-02	5.578e-01
##	2610	2.583e-01	3.506e-01	2.182e-08	1.970e-08	3.481e-05	4.113e-02
##	2611	5.163e-01	2.964e-01	2.576e-06	1.291e-05	3.333e-10	1.903e-15
##	2612	8.886e-01	2.885e-01	7.881e-02	4.037e-01	5.638e-03	3.128e-01
##	2613	1.037e-02	5.041e-01	4.845e-32	1.014e-35	9.269e-32	1.450e-33
##	2614	1.764e-01	4.894e-01	6.158e-07	2.994e-14	2.916e-05	2.430e-10
##	2700	1.474e-03	3.579e-01	3.487e-219	0.000e+00	8.124e-319	0.000e+00
##	2701	3.880e-01	9.562e-01	1.718e-29	1.546e-58	4.792e-60	1.516e-136
##	2702	8.159e-01	6.186e-01	8.186e-01	4.869e-01	7.787e-01	4.688e-02
##	2703	3.587e-02	5.628e-02	5.591e-35	9.406e-101	6.628e-98	1.178e-198
##	2704	3.514e-01	9.213e-01	2.416e-03	1.163e-04	3.914e-06	1.692e-09
##	2705	6.210e-15	1.830e-04	8.391e-19	2.194e-26	4.712e-57	2.023e-150
##	2706	1.027e-03	3.222e-01	1.227e-10	7.826e-36	9.415e-94	2.516e-257
##	2707	9.635e-01	4.318e-01	1.919e-02	4.803e-10	2.386e-13	1.384e-16
##	2708	1.401e-01	3.252e-02	1.615e-05	1.141e-12	2.581e-18	5.634e-40
##	2709	1.685e-03	2.831e-05	4.650e-01	9.941e-01	6.268e-06	5.839e-01
##	2710	4.002e-01	4.604e-02	7.335e-02	7.857e-02	1.928e-01	1.715e-02
##	2711	2.806e-01	6.962e-01	8.877e-03	2.633e-06	3.483e-13	2.155e-30
##	2712	9.564e-01	6.438e-01	6.181e-09	7.713e-20	1.199e-50	1.068e-115
##	2713	3.453e-01	7.509e-01	5.226e-06	1.043e-10	5.546e-10	2.231e-16
##	2714	3.361e-06	1.360e-02	7.879e-21	2.271e-37	3.223e-42	6.781e-59
##	2715	1.357e-03	9.246e-01	4.049e-01	1.171e-05	4.655e-08	1.719e-25
##	2716	1.703e-02	1.307e-03	9.598e-04	2.321e-06	1.326e-05	2.961e-16
##	2717	3.771e-02	1.802e-01	9.673e-07	9.535e-16	1.432e-21	8.459e-36
##	2718	6.531e-01	9.577e-01	2.440e-04	2.501e-07	2.841e-08	4.423e-13
##	2719	3.374e-04	7.808e-02	4.113e-15	5.964e-28	1.176e-27	3.127e-48
##	2720	2.456e-01	1.357e-02	2.818e-01	6.856e-05	2.080e-10	1.081e-33
##	2721	6.968e-02	4.266e-01	9.505e-02	1.558e-01	9.214e-03	5.775e-09
##	2722	6.644e-01	9.366e-01	3.263e-03	2.901e-04	1.189e-04	1.733e-09
##	2723	6.820e-01	4.713e-02	8.197e-19	7.283e-26	3.897e-33	2.935e-54
##	2724	1.531e-01	3.424e-01	7.011e-06	6.081e-13	5.249e-45	2.495e-131
##	2725	4.757e-01	8.236e-01	1.921e-04	5.427e-10	8.938e-13	1.454e-23
##	2726	8.526e-01	5.076e-01	3.726e-02	3.204e-03	5.960e-07	2.760e-15
##	2727	4.068e-01	5.951e-01	4.744e-01	4.427e-03	4.252e-05	1.181e-18
##	2728	6.280e-04	3.606e-03	5.342e-30	5.854e-56	1.534e-67	3.318e-137
##	2729	8.563e-04	1.316e-01	6.160e-24	4.653e-50	6.960e-86	9.707e-133
##	2730	7.493e-01	3.348e-01	1.915e-23	5.617e-38	1.136e-43	5.630e-113
##	2731	7.486e-01	2.129e-01	1.665e-11	5.830e-18	7.051e-20	5.854e-41
##	2732	6.818e-01	3.115e-01	6.030e-18	4.135e-27	4.843e-32	4.893e-64
##	2733	1.940e-02	3.140e-07	3.691e-02	1.347e-03	4.551e-04	8.567e-13
##	2734	8.420e-01	5.346e-01	4.501e-12	1.242e-07	4.519e-18	2.202e-51
##	2735	2.343e-04	1.020e-02	1.966e-20	1.099e-29	2.085e-52	5.906e-109
##	2736	6.599e-01	5.962e-01	4.053e-13	9.906e-36	3.738e-37	3.786e-50
##	2737	1.986e-02	6.328e-02	5.097e-01	1.788e-01	1.346e-01	5.907e-08
##	2738	6.628e-01	1.882e-06	3.248e-123	1.298e-218	1.139e-291	0.000e+00
##	2739	2.688e-03	3.703e-01	1.221e-40	1.533e-75	2.547e-88	1.560e-145
##	2740	7.014e-06	3.521e-04	2.306e-06	9.941e-09	7.537e-13	1.733e-39
##	2741	5.423e-02	3.686e-06	7.625e-20	1.201e-37	7.956e-47	1.570e-131

##	2742	6.680e-04	2.233e-05	1.090e-19	1.844e-36	6.066e-39	1.695e-39
##	2743	7.776e-01	9.772e-01	4.192e-07	1.112e-21	4.353e-40	2.150e-63
##	2744	2.528e-01	3.346e-01	3.929e-02	5.832e-01	2.927e-01	2.529e-01
##	2745	4.161e-01	9.735e-02	2.274e-03	3.345e-09	9.066e-13	1.425e-20
##	2746	1.784e-04	5.438e-04	4.311e-10	6.982e-30	1.594e-46	8.833e-146
##	2747	7.749e-01	7.742e-01	5.247e-01	7.917e-04	9.237e-05	3.569e-17
##	2748	4.440e-02	1.478e-06	1.549e-06	5.886e-10	5.239e-14	2.744e-27
##	2800	1.401e-01	3.103e-01	4.068e-28	4.743e-43	8.400e-46	1.249e-70
##	2801	4.856e-01	3.227e-02	7.557e-01	6.901e-01	6.902e-01	1.386e-02
##	2802	7.022e-01	2.194e-01	8.359e-07	6.562e-13	4.926e-14	1.124e-15
##	2803	1.687e-01	7.407e-02	3.940e-02	1.194e-02	6.319e-06	1.192e-05
##	2804	4.767e-01	2.145e-01	7.737e-03	3.962e-04	1.095e-04	1.019e-09
##	2805	2.882e-01	5.028e-01	9.893e-10	4.890e-14	5.484e-18	3.394e-28
##	2806	6.692e-01	3.399e-01	3.164e-04	9.771e-05	2.495e-05	4.863e-08
##	2807	1.286e-01	6.443e-01	7.525e-01	1.701e-01	1.514e-02	6.193e-02
##	2808	4.485e-01	4.617e-01	3.714e-08	4.964e-12	1.128e-17	7.094e-30
##	2809	5.324e-01	2.810e-02	1.861e-04	8.191e-08	7.091e-11	4.723e-18
##	2900	7.255e-07	4.213e-01	3.737e-50	5.125e-75	5.083e-54	5.358e-76
##	2901	5.747e-02	3.173e-01	7.214e-01	4.011e-02	4.233e-01	4.495e-02
##	2902	6.019e-01	2.228e-04	1.058e-17	6.124e-35	3.150e-47	2.576e-118
##	2903	6.147e-01	4.273e-01	7.518e-01	7.862e-02	8.532e-02	2.240e-02
##	2904	8.202e-01	4.136e-01	7.622e-01	7.811e-01	1.801e-01	1.721e-01
##	2905	2.160e-02	3.007e-01	3.677e-07	3.196e-03	7.755e-06	4.687e-05
##	2906	3.811e-01	3.000e-02	1.713e-01	9.989e-01	9.991e-01	1.466e-01
##	2907	9.925e-01	1.515e-01	2.697e-02	1.332e-08	3.320e-11	1.678e-23
##	2908	1.531e-01	7.871e-01	2.071e-03	1.379e-01	8.035e-01	2.624e-04
##	2909	7.311e-01	1.388e-01	1.078e-06	1.206e-10	1.820e-15	6.258e-14
##	2910	9.851e-01	7.091e-01	3.314e-01	6.210e-03	7.492e-02	3.003e-01
##	2911	3.514e-01	3.671e-02	1.008e-03	3.152e-08	1.820e-11	1.190e-15
##	2912	4.722e-01	8.987e-03	5.531e-01	1.728e-01	2.322e-01	4.927e-02
##	2913	1.272e-01	8.114e-01	1.514e-09	1.364e-14	3.130e-06	7.902e-09
##	2914	3.726e-02	5.599e-02	2.255e-01	1.597e-02	1.705e-05	4.780e-08
##	2915	NA	NA	NA	NA	NA	NA
##	2916	2.330e-01	1.873e-03	4.664e-28	1.066e-37	5.998e-39	2.233e-72
##	2917	9.691e-01	4.454e-01	6.136e-02	2.134e-05	4.484e-12	3.919e-12
##	2918	NA	NA	NA	NA	NA	NA
##	2919	5.910e-01	5.121e-01	2.355e-05	2.487e-06	3.301e-05	7.732e-03
##	2920	NA	NA	NA	NA	NA	NA
##	2921	7.163e-01	6.093e-02	3.389e-14	4.860e-15	8.350e-07	2.729e-11
##	2922	3.853e-01	2.170e-01	1.103e-03	1.170e-02	5.769e-04	5.399e-07
##	2923	7.876e-01	7.135e-01	6.313e-02	2.098e-01	9.775e-01	2.129e-01
##	3000	6.130e-01	2.276e-01	1.173e-03	3.148e-05	3.229e-06	1.520e-06
##	3001	NA	NA	NA	NA	NA	NA
##	3002	6.918e-01	5.877e-01	1.323e-04	3.367e-06	7.009e-08	4.522e-16
##	3003	6.134e-01	3.619e-01	2.275e-39	1.198e-119	3.706e-134	2.776e-243
##	3004	9.936e-01	8.793e-01	5.798e-38	1.278e-67	9.507e-77	3.062e-101
##	3005	9.152e-01	5.036e-01	1.237e-01	1.958e-02	1.073e-02	6.909e-05
##	3100	4.093e-01	4.890e-01	1.182e-25	6.058e-22	2.100e-28	3.569e-38
##	3101	3.098e-01	5.948e-01	2.074e-08	2.966e-13	2.281e-07	9.568e-16
##	3102	1.849e-01	8.335e-01	9.099e-16	1.741e-14	1.078e-10	1.369e-14

##	3103	3.896e-01	1.206e-01	2.765e-25	1.131e-32	2.466e-15	4.544e-15
##	3104	5.240e-01	6.209e-01	3.722e-09	3.510e-08	3.435e-09	4.021e-16
##	3105	7.335e-01	1.956e-01	2.302e-29	3.427e-29	1.784e-34	4.136e-40
##	3106	9.988e-01	4.586e-01	1.300e-20	3.293e-25	8.344e-14	2.869e-04
##	3107	2.403e-01	4.841e-01	1.417e-12	2.508e-17	1.396e-21	1.083e-39
##	3108	1.778e-02	2.541e-01	3.658e-02	1.188e-02	8.079e-04	9.273e-11
##	3109	5.460e-01	9.320e-01	2.596e-07	2.611e-11	1.387e-07	4.152e-04
##	3110	1.540e-01	3.908e-01	1.070e-07	1.198e-06	1.884e-10	5.557e-08
##	3200	2.965e-01	2.753e-01	3.175e-42	4.239e-52	3.350e-48	6.524e-45
##	3201	9.504e-01	3.800e-01	1.473e-02	1.379e-01	2.479e-01	3.830e-01
##	3202	3.172e-02	1.138e-02	1.347e-21	1.344e-23	2.884e-16	7.219e-30
##	3203	7.876e-02	3.416e-02	5.977e-80	3.696e-147	5.960e-147	1.750e-202
##	3204	2.392e-02	1.749e-01	1.665e-40	1.594e-62	6.840e-56	7.550e-86
##	3205	2.335e-01	3.725e-01	6.106e-20	8.341e-28	2.577e-28	5.190e-39
##	3206	6.135e-06	4.633e-02	1.783e-17	7.982e-29	3.216e-24	4.341e-28
##	3207	2.159e-03	5.157e-01	1.393e-09	7.620e-11	2.992e-09	1.924e-16
##	3300	3.065e-01	7.915e-01	7.655e-12	2.680e-11	5.652e-08	4.679e-03
##	3301	4.288e-01	5.077e-01	5.140e-09	4.758e-10	1.102e-08	3.810e-07
##	3302	8.726e-01	3.621e-02	1.869e-14	4.834e-18	5.308e-16	1.119e-11
##	3303	6.614e-03	2.616e-01	8.149e-17	1.447e-17	3.209e-05	1.737e-06
##	3304	1.455e-04	5.177e-04	4.340e-48	5.307e-46	9.639e-34	5.057e-41
##	3305	5.856e-04	1.381e-02	2.152e-48	7.743e-46	3.242e-11	2.537e-21
##	3306	1.148e-01	1.445e-01	1.910e-10	5.993e-17	3.818e-19	6.530e-15
##	3307	6.571e-02	3.779e-01	2.297e-01	3.786e-02	8.842e-02	7.387e-02
##	3308	5.857e-01	7.635e-01	1.821e-20	2.142e-17	2.710e-16	1.396e-07
##	3309	6.724e-01	8.722e-01	2.566e-48	6.156e-44	1.394e-13	1.295e-23
##	3310	4.975e-02	5.443e-01	1.214e-88	7.721e-88	4.520e-57	6.130e-37
##	3311	4.970e-01	2.103e-01	2.142e-01	2.061e-01	6.279e-01	4.333e-01
##	3312	9.767e-01	4.622e-03	1.315e-51	1.813e-30	5.038e-13	1.643e-11
##	3313	5.945e-02	3.772e-01	3.711e-16	1.442e-12	1.931e-08	1.577e-02
##	3314	8.999e-01	9.864e-02	4.926e-14	4.480e-11	4.024e-02	2.882e-01
##	3315	7.677e-03	7.383e-01	3.220e-16	2.786e-09	8.720e-03	2.156e-06
##	3316	4.988e-01	5.212e-01	1.210e-31	4.078e-13	1.076e-05	1.533e-04
##	3317	9.793e-01	1.152e-01	4.130e-04	7.424e-03	5.052e-07	1.487e-02
##	3318	2.736e-01	5.951e-01	1.870e-18	1.059e-07	7.226e-04	2.203e-01
##	3319	8.807e-01	7.037e-01	2.476e-03	3.999e-02	3.133e-05	3.050e-03
##	3320	9.407e-02	1.927e-02	9.802e-17	2.338e-09	3.376e-02	2.268e-04
##	3321	3.881e-01	4.809e-01	2.652e-07	2.965e-06	8.405e-03	1.707e-01
##	3322	5.671e-01	4.400e-01	5.658e-05	2.330e-10	9.664e-02	7.464e-05
##	3400	2.304e-06	7.371e-01	2.881e-14	1.033e-40	3.573e-62	7.673e-84
##	3401	7.701e-01	1.861e-01	5.044e-01	5.196e-01	7.492e-01	3.191e-01
##	3402	2.194e-01	8.999e-01	9.606e-01	1.072e-02	2.758e-02	1.025e-01
##	3403	3.168e-01	5.058e-01	6.552e-01	7.260e-01	7.736e-01	7.335e-01
##	3404	3.135e-01	5.259e-01	6.380e-03	6.495e-03	1.512e-03	1.107e-04
##	3500	1.070e-02	4.397e-01	2.827e-15	1.355e-22	2.538e-25	1.291e-37
##	3501	NA	NA	NA	NA	NA	NA
##	3502	NA	NA	NA	NA	NA	NA
##	3503	NA	NA	NA	NA	NA	NA
##	3504	4.918e-03	9.783e-02	2.243e-01	5.438e-02	4.313e-02	3.576e-04
##	3505	9.969e-02	4.082e-01	2.031e-02	9.306e-05	1.317e-03	1.947e-05

##	3506	7.045e-02	4.525e-01	2.599e-01	4.029e-01	8.506e-01	9.923e-01
##	3600	2.213e-01	3.110e-02	1.229e-03	2.697e-02	3.562e-03	1.433e-03
##	3601	5.558e-01	8.724e-01	2.469e-07	1.159e-06	6.018e-04	1.107e-03
##	3602	3.635e-01	1.305e-01	4.346e-02	7.523e-02	2.326e-01	9.725e-02
##	3603	9.634e-01	7.978e-01	1.644e-01	5.800e-03	1.109e-01	1.105e-02
##	3604	NA	NA	NA	NA	NA	NA
##	3605	2.673e-01	7.738e-03	2.191e-03	1.986e-04	4.061e-07	1.902e-04
##	3606	NA	NA	NA	NA	NA	NA
##	3607	1.341e-01	5.284e-01	1.387e-02	2.539e-02	1.663e-04	9.247e-03
##	3608	NA	NA	NA	NA	NA	NA
##	3609	3.568e-01	9.299e-01	9.609e-01	6.732e-01	3.080e-01	6.699e-02
##	3610	4.934e-01	2.148e-01	5.288e-03	2.266e-04	4.664e-03	9.173e-03
##	3611	5.653e-04	5.339e-01	3.960e-04	3.191e-07	1.784e-07	7.272e-08
##	3612	2.599e-06	3.941e-10	4.775e-28	2.420e-41	8.514e-42	4.882e-54
##	3613	6.286e-02	4.717e-01	3.863e-01	4.247e-02	1.662e-02	1.500e-02
##	3614	4.760e-05	9.674e-06	1.293e-09	4.276e-16	3.702e-17	4.122e-48
##	3615	NA	NA	NA	NA	NA	NA
##	3616	8.957e-01	4.853e-01	3.774e-06	9.039e-07	7.344e-06	1.566e-02
##		FFA2p	FLA2p	FFA3p	FLA4p		
##	1000	1.800e-20	4.220e-05	4.206e-14	4.953e-02		
##	1100	2.597e-03	8.791e-01	2.968e-03	7.976e-01		
##	1101	7.649e-01	7.680e-01	6.958e-01	7.009e-01		
##	1102	2.399e-05	5.462e-01	2.156e-05	2.953e-01		
##	1103	5.174e-15	3.849e-03	6.562e-18	5.522e-07		
##	1104	6.553e-01	1.355e-02	9.403e-01	1.583e-02		
##	1105	3.594e-01	9.582e-01	3.558e-01	8.249e-01		
##	1106	9.061e-05	2.046e-10	5.271e-02	7.802e-08		
##	1107	6.369e-01	2.478e-01	4.896e-01	2.161e-01		
##	1108	2.118e-01	8.547e-01	2.061e-01	9.335e-01		
##	1109	3.246e-01	2.716e-01	4.173e-01	3.387e-01		
##	1110	7.040e-05	7.506e-05	9.840e-04	8.473e-04		
##	1111	2.864e-01	2.936e-01	2.215e-01	2.266e-01		
##	1200	7.632e-01	8.451e-01	7.374e-01	8.406e-01		
##	1201	1.571e-02	1.612e-01	6.395e-02	9.216e-01		
##	1202	8.297e-01	4.688e-01	2.296e-01	1.590e-01		
##	1203	1.282e-03	7.668e-01	1.443e-06	2.663e-04		
##	1204	1.573e-01	8.671e-06	1.371e-02	1.727e-06		
##	1205	6.372e-01	6.858e-01	7.043e-01	7.738e-01		
##	1206	2.178e-01	3.420e-01	4.926e-01	9.955e-01		
##	1207	3.891e-14	4.781e-01	2.677e-29	7.521e-18		
##	1208	6.366e-01	8.109e-01	3.067e-01	3.619e-01		
##	1209	7.970e-01	6.250e-01	7.943e-01	6.708e-01		
##	1210	3.538e-02	7.323e-01	6.932e-03	5.102e-02		
##	1211	5.398e-01	5.248e-01	9.408e-01	8.579e-01		
##	1212	9.225e-02	5.336e-01	7.936e-04	3.355e-03		
##	1213	2.394e-01	8.238e-01	3.326e-03	6.027e-03		
##	1300	1.759e-03	3.338e-01	2.899e-03	5.947e-01		
##	1301	3.900e-02	4.130e-02	8.611e-02	2.154e-01		
##	1302	1.217e-01	1.642e-02	7.456e-02	1.076e-02		
##	1303	2.204e-02	4.272e-01	2.413e-02	4.832e-01		

```
## 1304 1.064e-01 9.704e-01 1.062e-01 9.037e-01
## 1305 2.848e-01 2.496e-03 5.106e-01 4.163e-03
## 1306 5.532e-01 4.475e-03 3.915e-01 4.151e-03
## 1307 5.753e-01 8.502e-01 5.637e-01 8.121e-01
## 1308 1.594e-06 1.853e-19 4.901e-10 1.443e-21
## 1309 3.381e-01 8.212e-01 3.254e-01 7.396e-01
## 1310 2.342e-01 4.690e-01 2.795e-01 5.698e-01
## 1311 4.918e-01 4.096e-02 6.680e-01 5.045e-02
## 1312 2.017e-01 2.107e-01 2.312e-01 2.414e-01
## 1313 4.470e-01 2.013e-01 4.866e-01 2.158e-01
## 1314 2.595e-01 1.769e-01 2.086e-01 1.443e-01
## 1315 2.226e-02 5.169e-01 2.403e-02 5.815e-01
## 1400 1.954e-01 9.107e-01 1.603e-01 4.976e-01
## 1401 5.130e-01 3.957e-01 1.849e-01 1.482e-01
## 1402 2.143e-01 5.752e-02 7.856e-03 2.521e-03
## 1403 4.896e-01 9.312e-01 3.920e-01 6.045e-01
## 1404 1.797e-01 7.924e-01 1.617e-01 5.838e-01
## 1405 7.554e-01 6.576e-01 5.388e-01 4.920e-01
## 1406 2.195e-03 3.290e-01 1.169e-04 8.017e-03
## 1407 1.378e-02 2.446e-01 1.502e-04 1.741e-03
## 1408 7.288e-01 5.853e-01 4.776e-01 4.176e-01
## 1409 2.403e-01 7.504e-01 2.592e-01 7.272e-01
## 1410 1.281e-01 2.432e-01 3.205e-01 7.478e-01
## 1500 7.506e-01 8.585e-02 5.472e-01 7.660e-02
## 1501 2.361e-01 1.313e-02 3.220e-04 8.209e-08
## 1502 7.383e-01 9.588e-01 7.453e-01 9.996e-01
## 1503 1.392e-02 5.669e-01 1.382e-02 5.587e-01
## 1504 3.876e-01 1.561e-02 9.615e-01 2.306e-02
## 1505 7.562e-02 7.694e-01 7.521e-02 7.538e-01
## 1506 2.603e-01 2.056e-01 1.772e-01 1.303e-01
## 1507 9.076e-01 5.298e-01 9.614e-01 5.527e-01
## 1508 1.627e-01 1.178e-01 2.382e-01 1.502e-01
## 1600 2.538e-01 2.663e-01 2.344e-01 2.457e-01
## 1601 1.144e-01 8.683e-01 1.084e-01 7.668e-01
## 1602 3.718e-05 3.890e-01 2.634e-05 2.414e-01
## 1603 3.500e-01 4.417e-01 3.138e-01 3.952e-01
## 1604 1.921e-01 8.825e-02 2.338e-01 1.060e-01
## 1605 9.619e-01 6.483e-03 9.509e-01 6.478e-03
## 1606 5.262e-02 9.889e-04 8.547e-02 1.507e-03
## 1607 2.441e-04 9.565e-01 2.471e-04 8.444e-01
## 1700 1.885e-02 2.172e-01 2.351e-03 2.117e-02
## 1701 2.689e-01 4.619e-02 9.943e-01 8.109e-02
## 1702 2.574e-01 4.219e-02 3.331e-02 6.938e-03
## 1703 9.440e-01 2.065e-01 7.834e-01 2.084e-01
## 1704 1.389e-01 4.994e-01 6.155e-02 1.761e-01
## 1705 5.927e-01 6.279e-02 7.064e-01 7.369e-02
## 1706 3.419e-01 1.718e-01 1.121e-01 6.725e-02
## 1707 6.367e-01 3.299e-01 8.301e-01 3.983e-01
## 1708 5.459e-01 7.955e-01 6.001e-01 9.602e-01
## 1709 2.730e-01 8.460e-01 2.979e-01 7.859e-01
```



```
## 1710 6.305e-01 4.396e-01 9.340e-01 5.544e-01
## 1711 6.662e-01 8.645e-01 6.976e-01 9.570e-01
## 1712 2.773e-02 7.631e-01 1.658e-02 2.652e-01
## 1800 4.330e-01 1.817e-01 7.471e-02 3.459e-02
## 1801          NA          NA          NA          NA
## 1802 1.916e-02 9.583e-01 4.919e-03 1.680e-01
## 1803 1.380e-01 3.619e-01 4.493e-02 9.826e-02
## 1804 5.249e-01 1.553e-01 9.543e-01 2.055e-01
## 1900 4.486e-01 7.537e-02 6.946e-02 1.746e-02
## 1901 6.743e-01 9.197e-01 6.580e-01 8.406e-01
## 1902 1.443e-02 4.941e-01 2.382e-02 9.368e-01
## 1903 7.655e-01 9.284e-01 6.958e-01 7.925e-01
## 1904 3.697e-01 6.110e-01 2.299e-01 3.251e-01
## 1905 6.081e-01 5.162e-02 8.660e-01 7.260e-02
## 1906 7.316e-01 6.440e-01 7.893e-01 6.789e-01
## 1907 2.146e-01 2.990e-01 3.136e-01 4.109e-01
## 1908 9.081e-01 5.323e-01 8.131e-01 5.189e-01
## 1909 4.212e-03 5.526e-01 5.848e-03 5.039e-01
## 1910 1.273e-02 8.687e-02 4.026e-02 2.446e-01
## 1911 1.434e-01 1.115e-01 3.136e-01 2.253e-01
## 1912 7.628e-01 9.981e-01 7.635e-01 9.579e-01
## 1913 8.537e-01 8.531e-02 7.347e-01 8.484e-02
## 2000 1.721e-01 4.923e-01 2.059e-02 4.299e-02
## 2001 6.726e-01 6.204e-01 8.899e-01 7.877e-01
## 2002 4.348e-01 2.705e-06 4.641e-05 9.783e-10
## 2003 8.989e-01 8.121e-01 7.763e-01 7.369e-01
## 2100 2.077e-01 3.722e-01 1.243e-01 2.306e-01
## 2101 1.849e-01 5.873e-01 2.545e-01 5.695e-01
## 2102 5.337e-01 6.121e-01 6.392e-01 7.295e-01
## 2103 8.264e-01 9.053e-01 7.869e-01 8.509e-01
## 2104 4.891e-01 6.671e-01 4.634e-01 6.303e-01
## 2105 8.801e-01 9.168e-02 2.570e-01 4.343e-02
## 2200 2.726e-03 3.807e-01 3.965e-04 3.378e-02
## 2201 3.562e-01 7.243e-01 3.860e-01 9.738e-01
## 2202 5.087e-01 6.276e-01 5.803e-01 7.817e-01
## 2203 3.841e-02 9.380e-01 3.039e-02 5.364e-01
## 2204 5.939e-01 6.438e-01 5.585e-01 6.042e-01
## 2205 6.592e-03 3.631e-01 1.331e-02 7.826e-01
## 2206 5.299e-01 7.405e-02 8.149e-01 9.500e-02
## 2207 4.496e-01 6.180e-01 3.233e-01 4.041e-01
## 2208 3.018e-01 3.789e-02 6.491e-01 6.030e-02
## 2209 7.005e-01 8.374e-01 7.523e-01 9.566e-01
## 2210 1.099e-01 9.172e-01 9.822e-02 6.325e-01
## 2211 2.641e-03 9.356e-01 2.198e-03 5.296e-01
## 2212 6.249e-01 7.221e-01 7.151e-01 8.160e-01
## 2213 5.946e-04 8.535e-01 3.057e-04 1.101e-01
## 2214 4.617e-01 3.012e-01 8.179e-02 6.818e-02
## 2215 2.120e-02 4.434e-01 3.888e-02 8.675e-01
## 2216 2.773e-04 2.981e-02 1.156e-02 8.974e-01
## 2300 2.243e-01 3.676e-01 3.453e-01 5.766e-01
```

```
## 2301 1.749e-01 3.496e-01 2.242e-02 3.806e-02
## 2302 8.816e-01 8.037e-01 8.985e-01 8.154e-01
## 2303 6.248e-02 5.239e-02 1.178e-01 9.712e-02
## 2304 7.321e-01 2.860e-01 7.799e-01 2.948e-01
## 2305 4.162e-01 1.528e-01 2.865e-01 1.181e-01
## 2306 6.078e-01 1.209e-01 9.890e-01 1.624e-01
## 2307 5.338e-01 1.319e-02 6.667e-01 1.551e-02
## 2308 1.804e-01 1.276e-01 5.936e-01 3.738e-01
## 2309 7.078e-04 2.036e-01 2.378e-03 6.944e-01
## 2310 7.561e-01 6.395e-01 7.861e-01 6.603e-01
## 2311 2.545e-01 4.419e-01 2.987e-01 5.432e-01
## 2312 6.716e-01 1.646e-02 9.295e-01 2.273e-02
## 2400 3.613e-01 2.645e-01 2.828e-01 2.125e-01
## 2401 1.588e-01 1.445e-01 7.574e-01 6.654e-01
## 2402 3.071e-02 9.538e-01 3.111e-02 7.986e-01
## 2403 4.227e-01 4.499e-01 4.419e-01 4.737e-01
## 2404 8.720e-01 5.691e-03 7.966e-01 5.657e-03
## 2405 3.486e-01 1.514e-02 4.685e-01 1.881e-02
## 2406 3.181e-03 1.749e-04 1.030e-03 5.888e-05
## 2500 8.169e-01 3.874e-01 7.045e-01 3.709e-01
## 2501 9.335e-01 1.372e-02 5.221e-02 4.640e-04
## 2502 7.963e-01 8.076e-01 7.723e-01 7.835e-01
## 2503 9.483e-01 2.655e-01 9.254e-01 2.658e-01
## 2504 3.379e-02 7.145e-01 3.191e-02 6.002e-01
## 2505 1.061e-01 2.073e-03 1.503e-01 2.747e-03
## 2506 2.731e-01 9.786e-01 2.755e-01 9.517e-01
## 2507 1.110e-01 1.394e-01 1.628e-01 1.872e-01
## 2508 3.591e-01 7.248e-01 3.519e-01 6.952e-01
## 2600 1.189e-02 1.820e-01 1.298e-05 2.943e-04
## 2601 4.572e-02 5.678e-01 4.026e-02 4.148e-01
## 2602 1.322e-01 1.603e-01 5.015e-01 6.288e-01
## 2603 7.077e-01 2.210e-01 6.794e-01 2.483e-01
## 2604 3.488e-01 4.435e-01 2.030e-01 2.471e-01
## 2605 2.173e-01 8.851e-01 2.008e-01 7.176e-01
## 2606 5.757e-01 2.790e-02 7.140e-01 2.610e-02
## 2607 6.597e-01 8.678e-01 5.242e-01 6.287e-01
## 2608 9.386e-01 9.002e-02 3.162e-02 1.883e-03
## 2609 5.802e-01 1.791e-01 7.693e-01 2.130e-01
## 2610 8.282e-01 7.531e-01 9.617e-01 8.050e-01
## 2611 7.979e-01 2.566e-01 5.824e-01 2.236e-01
## 2612 9.159e-01 2.571e-01 5.420e-01 2.041e-01
## 2613 6.491e-01 7.224e-01 5.473e-01 6.007e-01
## 2614 6.198e-02 7.024e-01 6.468e-02 6.835e-01
## 2700 8.352e-21 3.410e-17 2.894e-09 8.236e-06
## 2701 5.742e-04 2.131e-03 1.488e-02 4.999e-02
## 2702 9.491e-01 4.741e-01 8.691e-01 4.617e-01
## 2703 2.348e-05 4.408e-01 9.523e-07 7.491e-03
## 2704 6.451e-01 9.600e-01 6.410e-01 9.304e-01
## 2705 4.767e-19 2.330e-02 1.245e-20 2.804e-04
## 2706 6.796e-07 3.269e-01 6.632e-07 4.963e-01
```

## 2707 2.283e-01 1.225e-01 6.509e-01 2.632e-01  
## 2708 7.134e-01 8.069e-04 2.057e-01 4.046e-04  
## 2709 3.434e-03 3.992e-05 8.419e-01 1.423e-02  
## 2710 1.699e-01 3.297e-02 7.707e-02 1.559e-02  
## 2711 3.394e-01 7.485e-01 3.680e-01 9.405e-01  
## 2712 4.221e-02 6.294e-03 1.176e-01 1.465e-02  
## 2713 5.949e-03 3.976e-01 2.690e-03 1.317e-01  
## 2714 1.944e-11 3.798e-02 5.235e-12 5.644e-04  
## 2715 6.519e-06 9.665e-01 6.902e-06 6.560e-01  
## 2716 3.685e-01 1.506e-04 1.290e-01 8.866e-05  
## 2717 3.994e-03 6.169e-01 5.443e-03 7.801e-01  
## 2718 9.608e-01 5.870e-01 8.298e-01 5.689e-01  
## 2719 3.211e-10 1.918e-01 4.812e-13 2.642e-05  
## 2720 7.584e-03 1.014e-02 1.730e-02 2.239e-02  
## 2721 9.381e-03 6.244e-01 8.303e-03 4.758e-01  
## 2722 1.639e-01 8.345e-01 1.707e-01 9.291e-01  
## 2723 1.580e-03 4.707e-02 1.102e-03 2.822e-02  
## 2724 7.846e-04 9.140e-03 3.724e-03 3.778e-02  
## 2725 3.592e-01 4.579e-01 4.222e-01 5.398e-01  
## 2726 2.867e-01 6.011e-01 3.121e-01 6.794e-01  
## 2727 1.592e-01 9.190e-01 1.465e-01 6.851e-01  
## 2728 1.228e-08 1.871e-01 5.440e-09 3.747e-02  
## 2729 1.844e-05 3.588e-03 6.053e-04 1.083e-01  
## 2730 9.815e-02 8.439e-02 1.809e-01 1.535e-01  
## 2731 4.661e-01 2.357e-02 6.773e-01 3.107e-02  
## 2732 9.478e-01 1.169e-02 3.854e-01 1.011e-02  
## 2733 2.511e-02 9.047e-07 3.301e-04 1.502e-08  
## 2734 7.124e-02 4.886e-02 1.891e-01 1.234e-01  
## 2735 2.671e-06 5.740e-01 3.271e-06 4.100e-01  
## 2736 4.069e-03 8.785e-01 4.423e-03 8.454e-01  
## 2737 3.343e-02 4.844e-03 7.033e-02 1.038e-02  
## 2738 8.562e-07 3.458e-14 3.790e-03 1.558e-10  
## 2739 1.048e-10 1.101e-01 1.549e-09 5.877e-01  
## 2740 4.335e-08 1.152e-03 1.448e-09 3.685e-05  
## 2741 3.748e-01 8.662e-14 8.957e-03 2.384e-14  
## 2742 7.586e-03 2.258e-10 1.079e-07 2.497e-14  
## 2743 2.400e-01 3.679e-04 9.133e-01 8.893e-04  
## 2744 2.274e-01 2.684e-01 4.846e-01 8.050e-01  
## 2745 1.361e-01 1.368e-02 2.261e-01 2.184e-02  
## 2746 3.537e-04 2.054e-01 1.274e-04 5.933e-02  
## 2747 9.996e-01 9.109e-01 9.885e-01 9.105e-01  
## 2748 1.837e-02 4.469e-05 2.433e-03 6.348e-06  
## 2800 4.167e-01 1.252e-01 4.930e-01 1.445e-01  
## 2801 6.592e-01 3.505e-02 2.184e-01 1.727e-02  
## 2802 2.112e-01 8.033e-02 2.846e-01 1.059e-01  
## 2803 1.007e-01 9.491e-02 7.999e-02 7.422e-02  
## 2804 1.476e-01 3.188e-01 1.175e-01 2.449e-01  
## 2805 5.361e-01 9.461e-01 5.421e-01 9.592e-01  
## 2806 7.717e-01 2.095e-01 9.806e-01 2.229e-01  
## 2807 1.782e-01 9.497e-01 1.730e-01 9.041e-01

```
## 2808 8.837e-02 9.206e-01 8.802e-02 7.842e-01
## 2809 6.871e-02 9.611e-03 1.577e-01 2.092e-02
## 2900 4.331e-12 1.881e-01 1.868e-13 1.051e-03
## 2901 1.058e-02 1.869e-01 3.002e-02 8.504e-01
## 2902 1.524e-01 6.893e-23 1.133e-07 5.772e-30
## 2903 7.242e-01 2.649e-02 1.579e-01 1.230e-03
## 2904 6.465e-01 3.924e-01 4.959e-01 3.197e-01
## 2905 2.894e-02 4.718e-01 1.714e-02 2.450e-01
## 2906 3.148e-01 4.232e-02 7.917e-03 1.034e-03
## 2907 2.855e-01 2.832e-03 3.522e-05 5.360e-07
## 2908 6.282e-01 6.052e-01 2.170e-01 1.876e-01
## 2909 1.967e-01 5.796e-03 7.929e-01 7.701e-03
## 2910 8.564e-01 7.789e-01 9.972e-01 8.318e-01
## 2911 4.409e-03 4.139e-01 3.113e-04 1.576e-02
## 2912 6.180e-01 4.927e-03 6.534e-01 7.331e-03
## 2913 1.574e-04 1.872e-02 4.697e-04 9.440e-01
## 2914      NA      NA      NA      NA
## 2915      NA      NA      NA      NA
## 2916 7.923e-01 5.348e-09 6.067e-02 3.368e-09
## 2917 2.438e-01 6.925e-01 1.373e-01 3.114e-01
## 2918      NA      NA      NA      NA
## 2919 9.634e-01 6.349e-01 8.887e-01 6.421e-01
## 2920      NA      NA      NA      NA
## 2921 1.622e-01 1.774e-02 4.497e-03 6.594e-04
## 2922 7.129e-01 1.943e-01 5.714e-01 1.999e-01
## 2923 6.608e-01 8.687e-01 5.616e-01 6.685e-01
## 3000 4.781e-01 3.780e-01 5.359e-01 4.039e-01
## 3001      NA      NA      NA      NA
## 3002 8.101e-01 9.711e-01 8.102e-01 9.718e-01
## 3003 3.274e-01 4.925e-06 8.943e-01 1.370e-05
## 3004 6.038e-05 5.060e-01 8.786e-05 7.341e-01
## 3005 7.394e-01 3.008e-01 8.166e-01 3.161e-01
## 3100 1.962e-01 7.517e-01 1.647e-01 5.204e-01
## 3101 7.275e-01 3.873e-01 4.902e-01 2.969e-01
## 3102 5.277e-03 7.272e-01 3.929e-03 3.984e-01
## 3103 2.726e-01 1.367e-01 5.041e-01 2.235e-01
## 3104 5.856e-01 5.240e-01 5.264e-01 4.799e-01
## 3105 4.732e-02 2.513e-01 1.046e-01 4.924e-01
## 3106 7.551e-01 4.555e-01 9.429e-01 5.209e-01
## 3107 5.802e-01 4.420e-01 4.801e-01 3.806e-01
## 3108 9.369e-06 4.595e-02 1.070e-06 4.599e-03
## 3109 8.621e-01 5.727e-01 9.072e-01 5.880e-01
## 3110 2.333e-02 4.250e-01 2.632e-02 4.820e-01
## 3200 1.624e-02 2.635e-01 4.075e-02 7.389e-01
## 3201 7.313e-01 2.897e-01 3.881e-01 1.929e-01
## 3202 2.204e-01 1.987e-03 2.461e-02 3.861e-04
## 3203 1.841e-02 2.729e-04 2.112e-01 2.794e-03
## 3204 9.790e-01 7.475e-02 5.554e-01 6.855e-02
## 3205 3.672e-03 5.211e-01 5.322e-03 8.993e-01
## 3206 2.484e-09 8.469e-02 1.595e-09 2.300e-02
```

```
## 3207 9.435e-03 3.363e-01 8.688e-04 1.513e-02
## 3300 1.122e-01 9.834e-01 2.758e-02 1.197e-01
## 3301 1.102e-01 5.493e-01 9.776e-03 3.163e-02
## 3302 2.147e-01 2.823e-04 1.076e-02 1.438e-05
## 3303 1.139e-02 2.194e-01 8.346e-07 9.907e-06
## 3304 6.783e-07 3.045e-04 1.473e-17 3.866e-15
## 3305 3.158e-04 6.318e-02 5.942e-12 6.329e-10
## 3306 5.297e-03 5.145e-01 2.489e-04 1.181e-02
## 3307 2.501e-02 3.830e-01 4.983e-02 7.489e-01
## 3308 7.831e-01 9.380e-01 5.943e-01 6.462e-01
## 3309 8.265e-02 3.750e-01 1.341e-01 7.059e-01
## 3310 9.458e-05 8.903e-01 6.451e-08 1.417e-04
## 3311 3.820e-01 1.785e-01 8.818e-01 2.988e-01
## 3312 3.568e-01 7.387e-03 1.160e-05 5.238e-07
## 3313 5.352e-02 3.542e-01 1.027e-02 4.855e-02
## 3314 4.663e-01 2.603e-01 3.650e-03 2.539e-03
## 3315 1.000e-02 6.596e-01 6.099e-04 1.775e-02
## 3316 4.000e-01 5.149e-01 5.858e-01 8.269e-01
## 3317 6.545e-01 2.888e-01 7.198e-02 4.203e-02
## 3318 7.850e-01 5.613e-01 2.149e-01 1.810e-01
## 3319 6.054e-01 4.574e-01 8.952e-01 5.877e-01
## 3320 3.091e-01 3.571e-03 4.687e-12 1.172e-13
## 3321 2.727e-01 3.301e-01 3.905e-02 4.591e-02
## 3322 4.587e-01 4.918e-01 1.394e-01 1.516e-01
## 3400 3.224e-18 1.610e-01 4.158e-17 8.162e-02
## 3401 4.953e-01 3.238e-01 5.683e-01 3.994e-01
## 3402 2.225e-01 7.745e-01 2.211e-01 5.952e-01
## 3403 2.738e-01 5.634e-01 2.432e-01 4.968e-01
## 3404 3.228e-02 2.058e-01 1.030e-02 5.199e-02
## 3500 9.955e-05 2.932e-02 1.659e-03 4.015e-01
## 3501          NA          NA          NA          NA
## 3502          NA          NA          NA          NA
## 3503          NA          NA          NA          NA
## 3504 2.049e-03 6.646e-02 4.103e-04 1.335e-02
## 3505 4.636e-01 7.071e-01 5.253e-01 8.097e-01
## 3506 4.604e-02 4.626e-01 6.770e-02 6.272e-01
## 3600 2.609e-01 9.236e-02 7.881e-01 2.119e-01
## 3601 8.036e-01 8.349e-01 8.713e-01 9.318e-01
## 3602 2.965e-01 4.915e-02 8.642e-01 8.098e-02
## 3603 8.217e-01 8.467e-01 8.551e-01 9.335e-01
## 3604          NA          NA          NA          NA
## 3605 4.300e-01 1.583e-02 7.775e-01 3.257e-02
## 3606          NA          NA          NA          NA
## 3607 2.181e-01 4.518e-01 1.234e-01 2.204e-01
## 3608          NA          NA          NA          NA
## 3609 2.991e-01 7.699e-01 3.230e-01 8.557e-01
## 3610 9.959e-01 2.792e-01 8.813e-01 2.819e-01
## 3611 2.429e-05 7.318e-01 1.002e-05 1.388e-01
## 3612 2.135e-04 1.321e-14 3.917e-14 5.243e-23
## 3613 4.628e-02 4.418e-01 2.886e-02 1.374e-01
```

```
## 3614 1.667e-04 4.118e-12 1.468e-10 1.080e-16
## 3615          NA          NA          NA          NA
## 3616 5.575e-01 7.543e-01 4.696e-01 5.865e-01
```

```
print(RegStar)
```

```
##      FFA1p FLA1p 2p    3p    4p    5+p  FFA2p FLA2p FFA3p FLA4p
## 1000 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1100 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1101 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1102 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1103 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1104 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1105 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1106 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1107 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1108 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1109 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1110 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1111 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1200 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1201 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1202 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1203 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1204 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1205 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1206 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1207 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1208 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1209 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1210 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1211 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1212 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1213 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1300 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1301 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1302 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1303 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1304 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1305 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1306 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1307 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1308 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1309 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1310 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1311 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1312 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1313 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1314 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
## 1315 ""      ""      ""      ""      ""      ""      ""      ""      ""      ""
```

##	1400	" "	" "	"***"	"***"	"***"	"**"	" "	" "	" "	" "
##	1401	" "	" "	"***"	"***"	"***"	"**"	" "	" "	" "	" "
##	1402	" "	"*"	"**"	"**"	"**"	"***"	" "	" "	"**"	"**"
##	1403	" "	" "	"***"	"***"	"***"	" "	" "	" "	" "	" "
##	1404	" "	" "	"***"	"**"	"**"	" "	" "	" "	" "	" "
##	1405	" "	" "	"***"	"***"	"***"	"**"	" "	" "	" "	" "
##	1406	"*"	" "	"***"	"***"	"***"	"**"	"**"	" "	"***"	"**"
##	1407	"*"	" "	"***"	"***"	"***"	"***"	"**"	" "	"***"	"**"
##	1408	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1409	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	1410	" "	" "	"***"	"***"	"**"	"***"	" "	" "	" "	" "
##	1500	" "	"*"	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1501	"*"	" "	" "	"***"	"***"	"***"	" "	"**"	"***"	"***"
##	1502	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1503	"**"	" "	" "	" "	"**"	"**"	"**"	" "	"**"	" "
##	1504	" "	"*"	"**"	" "	" "	" "	" "	"**"	" "	"**"
##	1505	"*"	" "	" "	"**"	"**"	"***"	" "	" "	" "	" "
##	1506	" "	" "	"**"	"***"	" "	" "	" "	" "	" "	" "
##	1507	" "	" "	" "	"**"	"**"	"***"	" "	" "	" "	" "
##	1508	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1600	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1601	" "	" "	"***"	"***"	"**"	"***"	" "	" "	" "	" "
##	1602	"*"	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	" "
##	1603	" "	" "	" "	" "	"**"	"**"	" "	" "	" "	" "
##	1604	" "	" "	" "	" "	" "	"**"	" "	" "	" "	" "
##	1605	" "	"**"	" "	"**"	"***"	"***"	" "	"**"	" "	"**"
##	1606	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	" "	"***"
##	1607	"*"	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	" "
##	1700	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"**"	"**"
##	1701	" "	"*"	"**"	" "	" "	" "	" "	"**"	" "	" "
##	1702	" "	"*"	"***"	"***"	"***"	"***"	" "	"**"	"**"	"**"
##	1703	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1704	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1705	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1706	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1707	" "	" "	"**"	"**"	"**"	"***"	" "	" "	" "	" "
##	1708	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1709	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1710	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1711	" "	" "	"**"	"**"	"**"	"***"	" "	" "	" "	" "
##	1712	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"**"	" "
##	1800	" "	" "	" "	"**"	" "	"**"	" "	" "	" "	"**"
##	1801	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
##	1802	"*"	" "	" "	" "	" "	"**"	"**"	" "	"**"	" "
##	1803	" "	" "	"***"	"***"	"***"	" "	" "	" "	"**"	" "
##	1804	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1900	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	"**"
##	1901	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
##	1902	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"**"	" "
##	1903	" "	" "	" "	"**"	" "	"***"	" "	" "	" "	" "

## 1904	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 1905	" "	"*"	" "	"***"	"***"	"***"	" "	" "	" "	" "
## 1906	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 1907	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 1908	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 1909	"*"	" "	"***"	"***"	"**"	"***"	"**"	" "	"**"	" "
## 1910	" "	" "	"***"	"***"	"***"	"***"	"*"	" "	"*"	" "
## 1911	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 1912	"*"	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 1913	" "	" "	"*"	"*"	" "	"*"	" "	" "	" "	" "
## 2000	" "	" "	"***"	"***"	"**"	" "	" "	" "	"*"	"*"
## 2001	" "	" "	"***"	"***"	"**"	"**"	" "	" "	" "	" "
## 2002	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	"***"	"***"
## 2003	" "	" "	"***"	"***"	"***"	"*"	" "	" "	" "	" "
## 2100	" "	" "	"**"	" "	"*"	"*"	" "	" "	" "	" "
## 2101	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2102	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2103	"*"	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2104	" "	" "	"**"	"**"	"**"	"***"	" "	" "	" "	" "
## 2105	" "	"*"	"***"	"***"	"***"	"***"	" "	" "	" "	"*"
## 2200	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"***"	"*"
## 2201	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2202	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2203	"*"	" "	"***"	"***"	"***"	"*"	"*"	" "	"*"	" "
## 2204	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2205	"*"	" "	"***"	"***"	"***"	"***"	"**"	" "	"*"	" "
## 2206	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2207	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2208	" "	" "	"***"	"***"	"***"	"***"	" "	"*"	" "	" "
## 2209	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2210	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2211	" "	" "	"***"	"***"	"***"	"***"	"**"	" "	"**"	" "
## 2212	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2213	"*"	" "	"***"	"***"	"***"	"***"	"***"	" "	"***"	" "
## 2214	" "	" "	"***"	"**"	"*"	"***"	" "	" "	" "	" "
## 2215	" "	" "	"***"	"***"	"***"	"***"	"*"	" "	"*"	" "
## 2216	"**"	" "	"***"	"***"	"***"	"***"	"***"	"*"	"*"	" "
## 2300	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2301	" "	" "	" "	" "	" "	" "	" "	" "	"*"	"*"
## 2302	" "	" "	" "	" "	" "	"**"	" "	" "	" "	" "
## 2303	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2304	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2305	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2306	" "	" "	"***"	"***"	"**"	"***"	" "	" "	" "	" "
## 2307	" "	"*"	"***"	"***"	"***"	"***"	" "	"*"	" "	"*"
## 2308	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2309	"**"	" "	"***"	"***"	"***"	"***"	"***"	" "	"**"	" "
## 2310	" "	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2311	"*"	" "	"***"	"***"	"***"	"***"	" "	" "	" "	" "
## 2312	" "	" "	"***"	"***"	"***"	"***"	" "	"*"	" "	"*"



```
## 2400 "***" "" "****" "****" "****" "****" "" "" "" ""
## 2401 "" "" "" "" "" "" "" "" "" ""
## 2402 "" "" "****" "****" "****" "****" "*" "" "" ""
## 2403 "" "" "****" "****" "****" "****" "" "" "" ""
## 2404 "" "*" "****" "****" "****" "****" "" "*" "" "*"
## 2405 "" "" "" "" "****" "****" "" "*" "" "*"
## 2406 "***" "****" "****" "*" "****" "****" "****" "****" "****" "****"
## 2500 "*" "" "****" "****" "****" "****" "" "" "" ""
## 2501 "" "" "****" "****" "****" "****" "" "*" "" "****"
## 2502 "" "" "****" "****" "****" "****" "" "" "" ""
## 2503 "" "" "****" "****" "****" "****" "" "" "" ""
## 2504 "" "" "****" "****" "****" "****" "*" "" "*" ""
## 2505 "" "****" "****" "****" "****" "****" "" "*" "" "*"
## 2506 "" "" "****" "****" "****" "****" "" "" "" ""
## 2507 "" "" "****" "****" "****" "****" "" "" "" ""
## 2508 "" "" "****" "****" "****" "****" "" "" "" ""
## 2600 "****" "" "****" "****" "****" "****" "*" "" "****" "****"
## 2601 "*" "" "****" "****" "" "" "*" "" "*" ""
## 2602 "" "" "" "*" "" "" "" "" "" ""
## 2603 "" "" "*" "" "*" "****" "" "" "" ""
## 2604 "*" "" "****" "****" "****" "****" "" "" "" ""
## 2605 "" "" "****" "****" "****" "****" "" "" "" ""
## 2606 "" "****" "****" "****" "" "" "" "" "*" "" "*"
## 2607 "" "" "****" "" "****" "****" "" "" "" ""
## 2608 "" "" "****" "" "" "****" "" "" "" "*" "*"
## 2609 "" "" "****" "****" "" "" "" "" "" "" ""
## 2610 "" "" "****" "****" "****" "*" "" "" "" ""
## 2611 "" "" "****" "****" "****" "****" "" "" "" ""
## 2612 "" "" "" "" "*" "" "" "" "" ""
## 2613 "*" "" "****" "****" "****" "****" "" "" "" ""
## 2614 "" "" "****" "****" "****" "****" "" "" "" ""
## 2700 "***" "" "****" "****" "****" "****" "****" "****" "****" "****"
## 2701 "" "" "****" "****" "****" "****" "****" "****" "*" "*"
## 2702 "" "" "" "" "" "*" "" "" "" ""
## 2703 "*" "" "****" "****" "****" "****" "****" "" "****" "*"
## 2704 "" "" "****" "****" "****" "****" "" "" "" ""
## 2705 "****" "****" "****" "****" "****" "****" "****" "*" "****" "****"
## 2706 "***" "" "****" "****" "****" "****" "****" "" "****" ""
## 2707 "" "" "*" "****" "****" "****" "" "" "" ""
## 2708 "" "*" "****" "****" "****" "****" "" "****" "" "****"
## 2709 "***" "****" "" "" "****" "" "****" "****" "" "*"
## 2710 "" "*" "" "" "" "*" "" "*" "" "*"
## 2711 "" "" "****" "****" "****" "****" "" "" "" ""
## 2712 "" "" "****" "****" "****" "****" "*" "*" "" "*"
## 2713 "" "" "****" "****" "****" "****" "****" "" "****" ""
## 2714 "****" "*" "****" "****" "****" "****" "****" "*" "****" "****"
## 2715 "***" "" "" "****" "****" "****" "****" "" "****" ""
## 2716 "*" "****" "****" "****" "****" "****" "" "****" "" "****"
## 2717 "*" "" "****" "****" "****" "****" "*" "" "*" ""
## 2718 "" "" "****" "****" "****" "****" "" "" "" ""
```

## 2719 "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*"  
## 2720 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*"  
## 2721 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "\*\*\*\*" ""  
## 2722 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2723 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*"  
## 2724 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*"  
## 2725 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2726 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2727 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2728 "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*"  
## 2729 "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" ""  
## 2730 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2731 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2732 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2733 "" "" "\*\*\*\*" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*"  
## 2734 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2735 "\*\*\*\*" "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2736 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2737 "" "" "" "" "" "" "\*\*\*\*" "" "" "" ""  
## 2738 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*"  
## 2739 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2740 "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*"  
## 2741 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2742 "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*"  
## 2743 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2744 "" "" "" "" "" "" "" "" "" ""  
## 2745 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2746 "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2747 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2748 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2800 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2801 "" "" "" "" "" "" "" "" "" ""  
## 2802 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2803 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2804 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2805 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2806 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2807 "" "" "" "" "" "" "" "" "" ""  
## 2808 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2809 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2900 "\*\*\*\*" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2901 "" "" "" "" "" "" "" "" "" ""  
## 2902 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2903 "" "" "" "" "" "" "" "" "" ""  
## 2904 "" "" "" "" "" "" "" "" "" ""  
## 2905 "" "" "\*\*\*\*" "" "" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2906 "" "" "" "" "" "" "" "" "" ""  
## 2907 "" "" "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""  
## 2908 "" "" "" "" "" "" "\*\*\*\*" "" "" "" ""  
## 2909 "" "" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "\*\*\*\*" "" "" "" ""

## 2910	" "	" "	" "	"**"	" "	" "	" "	" "	" "	" "
## 2911	" "	"*"	"***"	"****"	"****"	"****"	"***"	" "	"****"	"*"
## 2912	" "	"***"	" "	" "	" "	"*"	" "	"**"	" "	"**"
## 2913	" "	" "	"****"	"****"	"****"	"****"	"****"	"*"	"****"	" "
## 2914	"*"	" "	" "	"*"	"****"	"****"	" "	" "	" "	" "
## 2915	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2916	" "	"**"	"****"	"****"	"****"	"****"	" "	"****"	" "	"****"
## 2917	" "	" "	" "	"****"	"****"	"****"	" "	" "	" "	" "
## 2918	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2919	" "	" "	"****"	"****"	"****"	"**"	" "	" "	" "	" "
## 2920	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 2921	" "	" "	"****"	"****"	"****"	"****"	" "	"*"	"**"	"****"
## 2922	" "	" "	"**"	"*"	"****"	"****"	" "	" "	" "	" "
## 2923	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 3000	" "	" "	"**"	"****"	"****"	"****"	" "	" "	" "	" "
## 3001	" "	" "	" "	" "	" "	" "	" "	" "	" "	" "
## 3002	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3003	" "	" "	"****"	"****"	"****"	"****"	" "	"****"	" "	"****"
## 3004	" "	" "	"****"	"****"	"****"	"****"	"****"	" "	"****"	" "
## 3005	" "	" "	" "	"*"	"*"	"****"	" "	" "	" "	" "
## 3100	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3101	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3102	" "	" "	"****"	"****"	"****"	"****"	"**"	" "	"**"	" "
## 3103	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3104	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3105	" "	" "	"****"	"****"	"****"	"****"	"*"	" "	" "	" "
## 3106	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3107	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3108	"*"	" "	"*"	"*"	"****"	"****"	"****"	"*"	"****"	"**"
## 3109	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3110	" "	" "	"****"	"****"	"****"	"****"	"*"	" "	"*"	" "
## 3200	" "	" "	"****"	"****"	"****"	"****"	"*"	" "	"*"	" "
## 3201	" "	" "	"*"	" "	" "	" "	" "	" "	" "	" "
## 3202	"*"	"*"	"****"	"****"	"****"	"****"	" "	"**"	"*"	"****"
## 3203	" "	"*"	"****"	"****"	"****"	"****"	"*"	"****"	" "	"**"
## 3204	"*"	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3205	" "	" "	"****"	"****"	"****"	"****"	"**"	" "	"**"	" "
## 3206	"****"	"*"	"****"	"****"	"****"	"****"	"****"	" "	"****"	"*"
## 3207	"**"	" "	"****"	"****"	"****"	"****"	"**"	" "	"****"	"*"
## 3300	" "	" "	"****"	"****"	"****"	"**"	" "	" "	"*"	" "
## 3301	" "	" "	"****"	"****"	"****"	"****"	" "	" "	"**"	"*"
## 3302	" "	"*"	"****"	"****"	"****"	"****"	" "	"****"	"*"	"****"
## 3303	"**"	" "	"****"	"****"	"****"	"****"	"*"	" "	"****"	"****"
## 3304	"****"	"****"	"****"	"****"	"****"	"****"	"****"	"****"	"****"	"****"
## 3305	"****"	"*"	"****"	"****"	"****"	"****"	"****"	" "	"****"	"****"
## 3306	" "	" "	"****"	"****"	"****"	"****"	"**"	" "	"****"	"*"
## 3307	" "	" "	" "	"*"	" "	" "	"*"	" "	"*"	" "
## 3308	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3309	" "	" "	"****"	"****"	"****"	"****"	" "	" "	" "	" "
## 3310	"*"	" "	"****"	"****"	"****"	"****"	"****"	" "	"****"	"****"

```

## 3311 "" "" "" "" "" "" "" "" "" ""
## 3312 "" "" "" "" "" "" "" "" "" "" ""
## 3313 "" "" "" "" "" "" "" "" "" "" ""
## 3314 "" "" "" "" "" "" "" "" "" "" ""
## 3315 "" "" "" "" "" "" "" "" "" "" ""
## 3316 "" "" "" "" "" "" "" "" "" "" ""
## 3317 "" "" "" "" "" "" "" "" "" "" ""
## 3318 "" "" "" "" "" "" "" "" "" "" ""
## 3319 "" "" "" "" "" "" "" "" "" "" ""
## 3320 "" "" "" "" "" "" "" "" "" "" ""
## 3321 "" "" "" "" "" "" "" "" "" "" ""
## 3322 "" "" "" "" "" "" "" "" "" "" ""
## 3400 "" "" "" "" "" "" "" "" "" "" ""
## 3401 "" "" "" "" "" "" "" "" "" "" ""
## 3402 "" "" "" "" "" "" "" "" "" "" ""
## 3403 "" "" "" "" "" "" "" "" "" "" ""
## 3404 "" "" "" "" "" "" "" "" "" "" ""
## 3500 "" "" "" "" "" "" "" "" "" "" ""
## 3501 "" "" "" "" "" "" "" "" "" "" ""
## 3502 "" "" "" "" "" "" "" "" "" "" ""
## 3503 "" "" "" "" "" "" "" "" "" "" ""
## 3504 "" "" "" "" "" "" "" "" "" "" ""
## 3505 "" "" "" "" "" "" "" "" "" "" ""
## 3506 "" "" "" "" "" "" "" "" "" "" ""
## 3600 "" "" "" "" "" "" "" "" "" "" ""
## 3601 "" "" "" "" "" "" "" "" "" "" ""
## 3602 "" "" "" "" "" "" "" "" "" "" ""
## 3603 "" "" "" "" "" "" "" "" "" "" ""
## 3604 "" "" "" "" "" "" "" "" "" "" ""
## 3605 "" "" "" "" "" "" "" "" "" "" ""
## 3606 "" "" "" "" "" "" "" "" "" "" ""
## 3607 "" "" "" "" "" "" "" "" "" "" ""
## 3608 "" "" "" "" "" "" "" "" "" "" ""
## 3609 "" "" "" "" "" "" "" "" "" "" ""
## 3610 "" "" "" "" "" "" "" "" "" "" ""
## 3611 "" "" "" "" "" "" "" "" "" "" ""
## 3612 "" "" "" "" "" "" "" "" "" "" ""
## 3613 "" "" "" "" "" "" "" "" "" "" ""
## 3614 "" "" "" "" "" "" "" "" "" "" ""
## 3615 "" "" "" "" "" "" "" "" "" "" ""
## 3616 "" "" "" "" "" "" "" "" "" "" ""

```

```
print("Gender-based average team sizes in 2018")
```

```
## [1] "Gender-based average team sizes in 2018"
```

```
print(GenderAnalysed)
```

```

##      CitationSet Female1st FemaleLast
## 1000         582         215         139
## 1100         661         299         217

```

## 1101	56	19	9
## 1102	139	58	37
## 1103	293	138	97
## 1104	228	89	54
## 1105	612	241	147
## 1106	197	93	77
## 1107	88	29	22
## 1108	13	5	1
## 1109	84	32	23
## 1110	201	84	54
## 1111	71	19	9
## 1200	111	56	55
## 1201	749	389	362
## 1202	1271	473	468
## 1203	505	272	263
## 1204	229	89	68
## 1205	26	8	9
## 1206	79	38	37
## 1207	313	114	111
## 1208	654	319	319
## 1209	34	14	11
## 1210	108	38	37
## 1211	834	258	254
## 1212	396	115	115
## 1213	556	294	308
## 1300	1019	441	320
## 1301	30	10	9
## 1302	59	33	22
## 1303	611	222	116
## 1304	245	84	50
## 1305	282	98	62
## 1306	268	132	75
## 1307	407	189	99
## 1308	126	58	27
## 1309	146	80	55
## 1310	140	65	43
## 1311	470	205	142
## 1312	680	308	174
## 1313	227	106	53
## 1314	333	142	76
## 1315	92	35	25
## 1400	276	99	92
## 1401	123	44	38
## 1402	177	57	47
## 1403	388	118	115
## 1404	58	18	14
## 1405	306	115	98
## 1406	186	68	71
## 1407	227	99	93
## 1408	568	209	183

## 1409	215	84	65
## 1410	30	10	10
## 1500	269	76	47
## 1501	28	9	4
## 1502	198	57	36
## 1503	377	106	53
## 1504	2	0	0
## 1505	75	19	11
## 1506	10	4	4
## 1507	78	17	8
## 1508	89	20	8
## 1600	855	239	148
## 1601	49	25	10
## 1602	160	58	25
## 1603	66	24	13
## 1604	179	57	38
## 1605	316	103	37
## 1606	475	121	69
## 1607	144	56	26
## 1700	161	28	25
## 1701	27	11	6
## 1702	42	11	8
## 1703	112	24	27
## 1704	58	12	10
## 1705	95	11	12
## 1706	317	59	65
## 1707	63	11	12
## 1708	55	11	6
## 1709	126	49	38
## 1710	188	44	39
## 1711	75	10	7
## 1712	220	39	28
## 1800	74	24	18
## 1801	0	0	0
## 1802	99	20	26
## 1803	153	35	42
## 1804	158	32	30
## 1900	150	49	39
## 1901	95	18	19
## 1902	219	68	37
## 1903	33	12	4
## 1904	164	66	45
## 1905	28	5	5
## 1906	115	28	20
## 1907	213	47	35
## 1908	106	25	19
## 1909	110	19	21
## 1910	101	35	26
## 1911	107	22	15
## 1912	137	25	17

## 1913	36	8	5
## 2000	119	29	34
## 2001	179	67	56
## 2002	807	238	216
## 2003	269	53	53
## 2100	152	39	28
## 2101	36	10	3
## 2102	202	48	26
## 2103	134	39	23
## 2104	93	31	19
## 2105	355	112	75
## 2200	160	33	25
## 2201	78	19	13
## 2202	94	16	7
## 2203	63	21	13
## 2204	305	97	60
## 2205	360	87	57
## 2206	24	2	2
## 2207	66	16	5
## 2208	185	23	10
## 2209	269	54	47
## 2210	552	91	52
## 2211	310	54	43
## 2212	54	8	9
## 2213	174	55	38
## 2214	44	20	19
## 2215	242	60	26
## 2216	94	37	35
## 2300	298	107	72
## 2301	175	75	86
## 2302	44	18	10
## 2303	360	142	100
## 2304	211	85	51
## 2305	153	47	37
## 2306	161	73	44
## 2307	134	60	46
## 2308	619	247	201
## 2309	255	92	76
## 2310	203	74	52
## 2311	135	49	36
## 2312	183	55	41
## 2400	220	98	55
## 2401	8	4	2
## 2402	68	30	17
## 2403	323	143	85
## 2404	224	124	74
## 2405	103	47	25
## 2406	101	61	34
## 2500	630	140	93
## 2501	26	9	10

## 2502	137	50	30
## 2503	159	44	21
## 2504	235	42	30
## 2505	142	43	24
## 2506	143	37	18
## 2507	94	24	17
## 2508	224	48	31
## 2600	256	33	32
## 2601	27	5	3
## 2602	93	22	17
## 2603	71	13	14
## 2604	372	62	43
## 2605	117	14	23
## 2606	67	13	6
## 2607	46	7	2
## 2608	55	8	5
## 2609	23	4	4
## 2610	105	7	5
## 2611	269	52	42
## 2612	21	4	3
## 2613	382	66	62
## 2614	130	19	17
## 2700	1288	646	491
## 2701	497	285	196
## 2702	53	31	17
## 2703	114	74	55
## 2704	32	18	11
## 2705	511	131	76
## 2706	154	41	37
## 2707	37	16	16
## 2708	116	69	66
## 2709	18	14	13
## 2710	27	14	10
## 2711	101	21	20
## 2712	229	109	66
## 2713	203	105	84
## 2714	91	60	34
## 2715	151	46	28
## 2716	174	101	75
## 2717	170	106	85
## 2718	147	66	53
## 2719	658	337	305
## 2720	122	57	36
## 2721	71	26	12
## 2722	45	25	6
## 2723	236	111	71
## 2724	117	50	23
## 2725	348	181	141
## 2726	144	68	49
## 2727	83	29	23



## 2728	538	237	137
## 2729	262	191	145
## 2730	386	180	121
## 2731	251	100	57
## 2732	673	173	139
## 2733	88	38	24
## 2734	284	140	122
## 2735	444	269	192
## 2736	353	180	119
## 2737	236	90	52
## 2738	995	615	467
## 2739	970	571	476
## 2740	265	116	73
## 2741	380	143	95
## 2742	202	116	99
## 2743	90	56	43
## 2744	0	0	0
## 2745	132	66	46
## 2746	639	151	99
## 2747	59	29	12
## 2748	134	43	25
## 2800	345	146	86
## 2801	30	12	8
## 2802	149	68	61
## 2803	119	69	47
## 2804	201	89	67
## 2805	308	141	96
## 2806	35	20	21
## 2807	22	10	7
## 2808	271	146	89
## 2809	140	59	31
## 2900	280	194	190
## 2901	34	17	17
## 2902	121	77	62
## 2903	4	3	3
## 2904	11	7	5
## 2905	157	114	107
## 2906	30	12	8
## 2907	19	8	6
## 2908	20	14	13
## 2909	68	53	43
## 2910	95	42	38
## 2911	41	12	15
## 2912	14	10	13
## 2913	97	88	78
## 2914	64	42	48
## 2915	3	3	0
## 2916	211	139	90
## 2917	46	28	31
## 2918	0	0	0

## 2919	42	30	29
## 2920	0	0	0
## 2921	148	96	74
## 2922	20	11	12
## 2923	24	12	10
## 3000	45	19	16
## 3001	10	2	4
## 3002	193	67	28
## 3003	215	86	59
## 3004	276	125	68
## 3005	109	45	36
## 3100	514	111	85
## 3101	160	17	15
## 3102	76	16	11
## 3103	189	37	30
## 3104	602	118	65
## 3105	187	34	23
## 3106	110	20	10
## 3107	352	55	41
## 3108	34	12	7
## 3109	134	11	10
## 3110	62	19	8
## 3200	384	200	172
## 3201	42	24	21
## 3202	411	228	188
## 3203	504	333	259
## 3204	563	375	313
## 3205	472	243	188
## 3206	138	75	54
## 3207	459	219	205
## 3300	380	169	165
## 3301	442	220	209
## 3302	233	83	75
## 3303	440	182	164
## 3304	1657	934	841
## 3305	1226	460	434
## 3306	623	382	357
## 3307	78	40	19
## 3308	1168	494	512
## 3309	191	72	86
## 3310	595	327	311
## 3311	136	48	52
## 3312	2304	978	938
## 3313	149	48	37
## 3314	507	230	222
## 3315	502	242	253
## 3316	1255	542	548
## 3317	108	67	62
## 3318	241	176	172
## 3319	124	75	67

## 3320	943	315	320
## 3321	228	98	91
## 3322	207	66	76
## 3400	266	162	120
## 3401	10	7	4
## 3402	10	7	9
## 3403	34	24	17
## 3404	14	5	5
## 3500	92	43	25
## 3501	4	1	1
## 3502	0	0	0
## 3503	0	0	0
## 3504	32	13	4
## 3505	15	12	8
## 3506	5	1	2
## 3600	25	14	16
## 3601	42	24	20
## 3602	15	8	8
## 3603	8	2	3
## 3604	6	5	2
## 3605	42	13	15
## 3606	0	0	0
## 3607	16	4	3
## 3608	4	3	3
## 3609	13	11	8
## 3610	20	9	5
## 3611	43	23	25
## 3612	417	157	130
## 3613	5	2	0
## 3614	71	26	21
## 3615	0	0	0
## 3616	83	55	45

```

write.table(BasicStats, file=paste("D:\\Downloads\\basicStats", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3b.txt", sep=""))
write.table(BasicStats2, file=paste("D:\\Downloads\\basicStats2", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3b.txt", sep=""))
write.table(GenderTeamSize, file=paste("D:\\Downloads\\GenderTeamSize",
Country, MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3b.txt", sep=""))
write.table(RegCoef, file=paste("D:\\Downloads\\RegCoef", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3b.txt", sep=""))
write.table(RegP, file=paste("D:\\Downloads\\RegP", Country,
MaxCountriesAllowed, " ", FirstYearForCitationAnalysis, "-",
LastYearForCitationAnalysis, "var3b.txt", sep=""))
write.table(RegStar, file=paste("D:\\Downloads\\RegStar", Country,

```

```
MaxCountriesAllowed," ", FirstYearForCitationAnalysis, "-",  
LastYearForCitationAnalysis, "var3b.txt", sep="")
```